

Л.Н. Волков,
М.С. Немировский,
Ю.С. Шинаков

СИСТЕМЫ ЦИФРОВОЙ РАДИОСВЯЗИ

Базовые методы и характеристики

Рекомендовано УМО по образованию в области телекоммуникаций в качестве учебного пособия для студентов высших учебных заведений, обучающихся по специальностям 201000 «Многоканальные телекоммуникационные системы»; 201100 «Радиосвязь, радиовещание и телевидение»; 201200 «Сети связи с подвижными объектами»

ЭКТРЕНДЗ

Москва, 2005

УДК 621.391.037.372

ББК 82.841

Волков Л.Н., Немировский М.С., Шинаков Ю.С.

Системы цифровой радиосвязи: базовые методы и характеристики: Учеб. пособие. — М.: Эко-Трендз, 2005. — 392 с.: ил.

ISBN 5-88405-071-2

Изложены теоретические основы построения систем цифровой радиосвязи, включая теорию информации, теорию потенциальной помехоустойчивости и основы помехоустойчивого кодирования. Введены и проанализированы количественные показатели энергетической и полосной эффективности систем связи, описаны пути совершенствования сигналов по этим показателям. Рассмотрены вопросы построения и использования полосно-эффективных и широкополосных сигналов, пути и методы многоканальной передачи сообщений при временном уплотнении. Отдельные разделы посвящены проблемам работы систем цифровой радиосвязи по многолучевым каналам и построению устройств синхронизации этих систем. Заключительная глава в качестве иллюстрации использования изложенных методов содержит краткий обзор вопросов построения современных перспективных систем наземной и спутниковой связи, в первую очередь, сотовых.

Для студентов вузов, обучающихся по направлениям 6504 «Телекоммуникации» и 6542 «Радиотехника». Может быть полезна инженерам, занимающимся разработкой и эксплуатацией современных цифровых систем связи.

ББК 82.841

ISBN 5-88405-071-2

© Авторы, 2005

**Федеральная целевая программа «Культура России»
(подпрограмма «Поддержка полиграфии и книгоиздания России»)**

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ	8
ВВЕДЕНИЕ	9
Глава 1. РАДИОВОЛНЫ, СИГНАЛЫ И ШУМЫ	11
1.1. Электромагнитные волны. Прием и излучение	11
1.1.1. Электромагнитные волны	11
1.1.2. Излучение и прием электромагнитных волн	13
1.1.3. Энергетический расчет радиолиний	16
1.2. Элементы теории вероятностей	18
1.2.1. Алгебра случайных событий	18
1.2.2. Случайные величины и распределения вероятностей	20
1.2.3. Случайные процессы	24
1.3. Видео- и радиосигналы. Цифровая модуляция	27
1.3.1. Видео- и радиосигналы	27
1.3.2. Цифровые представления видеосигналов	30
1.3.3. Цифровые представления радиосигналов	35
1.3.4. Базовые методы цифровой модуляции	37
Глава 2. ТЕОРИЯ ИНФОРМАЦИИ И ДОСТИЖИМАЯ ЭФФЕКТИВНОСТЬ СИСТЕМ ЦИФРОВОЙ РАДИОСВЯЗИ	41
2.1. Мера количества информации и производительность источника	41
2.1.1. Мера информации	42
2.1.2. Производительность источника информации	44
2.2. Информация при детерминированных преобразованиях сообщений и основная теорема для канала без шумов	45
2.2.1. Детерминированные преобразования сообщений	45
2.2.2. Основная теорема для канала без шумов	47
2.3. Информация при недетерминированных преобразованиях сигналов. Пропускная способность канала с шумами	50
2.3.1. Информация при недетерминированных преобразованиях	50
2.3.2. Канал связи с шумами и основная теорема для него	51
2.4. Предельная энергетическая и спектральная эффективность типовой системы связи	53
2.4.1. Пропускная способность типового канала связи	53
2.4.2. Достижимые показатели эффективности систем связи	55
Глава 3. ИДЕАЛЬНЫЙ ПРИЕМ ЦИФРОВЫХ СИГНАЛОВ	58
3.1. Постановка задачи идеального приема при цифровой передаче сообщений	58
3.1.1. Структура рассматриваемых радиолиний	58
3.1.2. Постановка задачи идеального приема	60
3.2. Когерентный идеальный прием	61
3.2.1. Структура идеального приемника	61
3.2.2. Потенциальная помехоустойчивость при двоичной передаче	65
3.2.3. Потенциальная помехоустойчивость при недвоичной передаче	69
3.2.4. Потенциальная помехоустойчивость при ортогональных ансамблях сигналов	71

3.3. Некогерентный прием	72
3.3.1. Сигналы со случайной начальной фазой и структура идеального некогерентного приемника.....	72
3.3.2. Помехоустойчивость некогерентного приема	75

Глава 4. ПОМЕХОУСТОЙЧИВОЕ КОДИРОВАНИЕ В СИСТЕМАХ ЦИФРОВОЙ РАДИОСВЯЗИ..... 78

4.1. Общие принципы помехоустойчивого кодирования и декодирования	78
4.2. Классификация помехоустойчивых кодов	81
4.2.1. Блочные коды	82
4.2.2. Древовидные коды	83
4.3. Кодирование методом укрупнения алфавита.....	85
4.3.1. Оценки помехоустойчивости при когерентном приеме	85
4.3.2. Оценки помехоустойчивости при некогерентном приеме	88
4.4. Циклические коды и методы их декодирования.....	89
4.4.1. Определения основных алгебраических структур	90
4.4.2. Циклические коды.....	91
4.4.3. Многотактные линейные фильтры	95
4.4.4. Наиболее распространенные циклические коды	98
4.4.5. Декодирование циклических кодов	102
4.4.6. Эффективность кодирования блоковыми кодами	107
4.5. Сверточные коды и методы их декодирования.....	111
4.5.1. Структура сверточных кодов	111
4.5.2. Алгоритм декодирования Витерби	116
4.5.3. Алгоритм последовательного декодирования	122
4.6. Каскадные коды	126
4.6.1. Основные принципы каскадного кодирования.....	126
4.6.2. Помехоустойчивость каскадных кодов	128
4.7. Турбокоды и их декодирование	129
4.7.1. Сверточные турбокоды	130
4.7.2. Блочные турбокоды	134

Глава 5. СПЕКТРАЛЬНО-ЭФФЕКТИВНАЯ ЦИФРОВАЯ МОДУЛЯЦИЯ И СИГНАЛЬНОЕ КОДИРОВАНИЕ..... 140

5.1. Спектрально-эффективные ансамбли сигналов.....	140
5.1.1. Введение.....	140
5.1.2. Построение полосно-эффективных сигналов	144
5.2. Общие данные о сигнальном кодировании	147
5.3. СКК на базе решетчатых сигнальных кодов.....	150
5.4. СКК на базе обобщенных каскадных кодов.....	156

Глава 6. МНОГОСТАНЦИОННЫЙ ДОСТУП И МНОГОКАНАЛЬНАЯ СВЯЗЬ..... 160

6.1. Общая теория и классификация методов линейного разделения сигналов	160
6.1.1. Постановка задачи.....	160

6.1.2. Условия разделимости сигналов	162
6.1.3. Классификация ортогональных сигналов и способов разделения	164
6.1.4. Многостанционный доступ при сотовой связи.....	171
6.2. Эффективность использования полосы рабочих частот в системах с разделением по частоте	173
6.2.1. Внеполосные излучения при базовых методах манипуляции.....	173
6.2.2. Подавление внеполосных излучений при ФМ путем сглаживания скачков фазы	177
6.2.3. Манипуляция с минимальным сдвигом	179
6.2.4. Гауссовская манипуляция с минимальным сдвигом.....	185
6.2.5. Подавление внеполосных излучений сигналов ФМ путем их фильтрации в передатчике	191
6.2.6. Частотное разделение сигналов при использовании общего передающего тракта для многоканальной передачи	195
6.3. Системы с временным разделением.....	201
6.3.1. Общие сведения.....	201
6.3.2. Цикловая синхронизация и скремблирование	204
6.3.3. Алгоритмы асинхронного ввода	208
Глава 7. ШИРОКОПОЛОСНЫЕ СИГНАЛЫ В СИСТЕМАХ СВЯЗИ.....	213
7.1. Методы широкополосной передачи.....	213
7.2. Шумоподобные сигналы и псевдослучайные последовательности.....	215
7.2.1. Последовательности Баркера	216
7.2.2. М-последовательности.....	216
7.2.3. Ансамбли псевдослучайных последовательностей.....	218
7.3. Некоторые особенности использования ШПС в системах связи	223
7.3.1. Эффективность использования отведенной полосы частот	223
7.3.2. Защита от замираний.....	226
7.3.3. Электромагнитная совместимость.....	226
7.3.4. Эффективность использования мощности бортового передатчика	226
7.4. Псевдослучайная перестройка рабочей частоты	227
7.5. Помехозащищенность широкополосных систем связи.....	228
7.5.1. Анализ помехозащищенности при ППРЧ	229
7.5.2. Эффективность и особенности кодирования при ППРЧ	233
7.5.3. Анализ помехозащищенности при ШПС	236
Глава 8. ВЕРОЯТНОСТНЫЕ МОДЕЛИ МНОГОЛУЧЕВЫХ РАДИОКАНАЛОВ	237
8.1. Многолучевой радиоканал.....	237
8.2. Мощность сигнала на выходе многолучевого канала	244
8.3. Числовые характеристики многолучевых каналов.....	249
8.3.1. Рассеяние во времени.....	249
8.3.2. Полоса когерентности по частоте	251
8.3.3. Рассеяние по частоте	252
8.3.4. Время когерентности	253

8.4.	Вероятностные модели сигналов на выходе многолучевых каналов	254
8.4.1.	Модель Кларка гладких замираний	254
8.4.2.	Доплеровский спектр в модели Кларка	256
8.4.3.	Имитационное моделирование многолучевого канала с рассеянием по частоте и по времени.....	258
8.5.	Модели каналов для испытаний оборудования некоторых радиосистем	261
8.5.1.	Рекомендации стандарта GSM-900 для наземных сетей	261
8.5.2.	Рекомендации стандарта GMR для спутниковых систем	265
8.5.3.	Модели каналов некоторых широкополосных систем третьего поколения.....	271

Глава 9.	МЕТОДЫ ОБЕСПЕЧЕНИЯ ЭФФЕКТИВНОЙ РАБОТЫ СИСТЕМ РАДИОСВЯЗИ ПО МНОГОЛУЧЕВЫМ КАНАЛАМ	275
9.1.	Классификация искажений сигналов, вызванных многолучевостью, и методов борьбы с ними	275
9.1.1.	Классификация искажений	275
9.1.2.	Классификация методов приема сигналов в многолучевых каналах	277
9.2.	Медленные дружные замирания и разнесенный прием	280
9.2.1.	Предпосылки и исходные соотношения	280
9.2.2.	Помехоустойчивость при отсутствии разнесения	283
9.2.3.	Кодирование с разнесением во времени	285
9.2.4.	Другие варианты разнесенного приема	289
9.3.	Когерентный прием цепочек элементарных сигналов в многолучевом канале.....	293
9.3.1.	Модель сигнала основной полосы	293
9.3.2.	Оценка максимального правдоподобия последовательности информационных символов.....	294
9.3.3.	Пример задачи динамического программирования	296
9.3.4.	Алгоритм Витерби.....	297
9.3.5.	Пример реализации алгоритма Витерби	302
9.3.6.	Сигналы с многопозиционной модуляцией	304
9.4.	Когерентный прием сигналов с информационной связью в многолучевом канале.....	306
9.4.1.	Решетчатая диаграмма сигнала гауссовской манипуляции с минимальным сдвигом (ГММС)	306
9.4.2.	Алгоритм Витерби демодуляции сигнала ГММС	309
9.4.3.	Эквалайзер Витерби	316
9.5.	Ортогональное частотное разделение с мультиплексированием	322
9.5.1.	Концепция ортогонального частотного разделения	322
9.5.2.	Формирование OFDM-радиосигнала	324
9.5.3.	Защитный временной интервал и циклическое продолжение.....	326
9.5.4.	Уменьшение внеполосных излучений	328
9.5.5.	Когерентная демодуляция OFDM-сигнала	332

Глава 10. СИНХРОНИЗАЦИЯ В СИСТЕМАХ ЦИФРОВОЙ РАДИОСВЯЗИ	334
10.1. Общие данные о системах автоматического регулирования, используемых для синхронизации	334
10.1.1. Общие сведения	334
10.1.2. Режимы работы САР и их анализ	336
10.1.3. Системы фазовой автоподстройки.....	339
10.2. Системы синхронизации по несущей и тактам при фазовой манипуляции	340
10.2.1. Цифровые представления входных данных систем синхронизации	340
10.2.2. Тактовая синхронизация	341
10.2.3. Синхронизация по несущей.....	343
10.2.4. Организация процессов поиска	345
10.2.5. Системы поиска и синхронизации по задержке для широкополосного сигнала	349
10.2.6. Система синхронизации пакетов.....	351
Глава 11. НЕКОТОРЫЕ ПЕРСПЕКТИВНЫЕ СИСТЕМЫ РАДИОСВЯЗИ.....	354
11.1. Спутниковые системы фиксированной службы радиосвязи и вещания	354
11.1.1. Искусственные спутники Земли в качестве ретрансляторов для систем связи	354
11.1.2. Диапазоны частот и зоны обслуживания	356
11.1.3. Структура и характеристики полезной нагрузки.....	357
11.1.4. Многостанционный доступ	358
11.1.5. Структура и характеристики земных средств.....	359
11.1.6. Тенденции развития и совершенствования.....	360
11.1.7. Энергетический расчет спутниковых радиолиний.....	362
11.2. Сотовые наземные системы связи.....	364
11.2.1. Общие сведения	365
11.2.2. GSM: частотный план и многостанционный доступ.....	367
11.2.3. GSM: радиointерфейсы и характеристики оборудования	368
11.2.4. CDMA. Общая характеристика	369
11.2.5. CDMA: формирование широкополосных сигналов	371
11.2.6. CDMA: частотный план и параметры аппаратуры.....	373
11.3. Сотовые спутниковые системы подвижной связи.....	374
11.3.1. Принципы построения и показатели эффективности.....	374
11.3.2. Глобальные и региональные системы	376
11.3.3. Система «Globalstar»	376
11.3.4. Система «Iridium».....	379
11.3.5. Система «Thuraya»	382
Список сокращений.....	386
Литература	388

ПРЕДИСЛОВИЕ

Большая часть материала настоящей книги посвящена изложению теоретических основ построения цифровых систем радиосвязи, начиная с теории потенциальной помехоустойчивости и теории информации. Далее в книге рассматриваются вопросы помехоустойчивого кодирования и многостанционного доступа применительно к различным системам цифровой радиосвязи. Отдельные главы посвящены вопросам использования в таких системах широкополосных и полосно-эффективных сигналов, а также методам повышения устойчивости работы систем радиосвязи по многолучевым каналам. Поскольку характеристики современных систем связи в значительной мере зависят от эффективности работы разнообразных систем синхронизации, охватывающих как систему в целом, так и отдельные ее звенья, в книге специальный раздел посвящен вопросам построения соответствующих систем автоматического управления.

Заключительный раздел книги призван проиллюстрировать применение изложенных выше теории и методов при построении современных систем связи. При этом в качестве примера рассматриваются системы фиксированной спутниковой связи и сотовые системы подвижной связи как наземные, так и спутниковые.

Для того чтобы облегчить ознакомление с основным материалом книги, ему предпослана вводная глава, содержащая сводку использованных сведений по вопросам распространения радиоволн, теории сигналов, включая вопросы аналого-цифрового преобразования, и теории вероятностей и случайных процессов.

Материал книги составлен на основе ряда лекционных курсов по направлению «радиотехнические системы», читавшихся авторами в течение многих лет студентам Московского технического университета связи и информатики (МТУСИ).

Следует отметить некоторые основные принципы, которыми руководствовались авторы как при работе над лекционными курсами, так и, как следствие, при отборе материала для данной книги. *Во-первых*, авторы не стремились к формально строгому изложению материала, обращая основное внимание на изложение основных идей, лежащих в основе современной радиосвязи. *Во-вторых*, авторы ставили своей целью расширить традиционное содержание рассматриваемых курсов за счет включения новых направлений и методов, возникших в системах радиосвязи в последние годы. Наконец, авторы считали целесообразным сочетать изучение студентами новых материалов с повторением основных положений ранее изученных курсов с тем, чтобы эти положения вошли бы, в конце концов, в тот «сухой остаток» знаний, который должен быть следствием обучения в высшей школе. Последним и объясняется включение в работу достаточно обширной вводной главы.

Существенное влияние на содержание предлагаемой вниманию читателей книги оказала и многолетняя деятельность авторов по созданию современных систем и аппаратуры связи. Это обстоятельство в значительной мере определило содержание книги и методику изложения материалов в ней. Авторы полагают поэтому, что книга будет полезной инженерам, занятым разработкой и эксплуатацией современного связного оборудования, а также всем тем, кто хотел бы изучить принципы его функционирования, по-существу, избегая излишне сложных формальных построений.

Главы 4 и 7 и пункт 11.3.5 написаны Л.Н. Волковым, главы 1–3, 5, 6 (кроме 6.2.4), 10, 11 (кроме 11.3.5) и § 9.1 и 9.2 — М.С. Немировским, а главы 8 и 9 (кроме § 9.1 и 9.2), а также пункт 6.2.4 — Ю.С. Шинаковым. Разработка общего плана книги, отбор материала и научное редактирование рукописи осуществлялись авторами совместно.

ВВЕДЕНИЕ

В соответствии с общепринятой терминологией под *радиосвязью* понимают связь, осуществляемую с помощью радиоволн. При этом к радиоволнам относят электромагнитные волны, распространяющиеся в пространстве без искусственного волновода, частоты которых ограничены 3000 ГГц. Указанная граница, хотя и является условной, тем не менее подчеркивает ограниченность частотного ресурса в системах радиосвязи.

Согласно «Регламенту радиосвязи», утвержденному Всемирной административной конференции радиосвязи в 1979 г., системы связи подразделяются на наземные и космические. *Космической* считается «...любая радиосвязь, при которой используется одна или несколько космических станций или один, или несколько отражающих спутников, или другие космические объекты». Все системы радиосвязи, кроме космических и радиоастрономии, отнесены к наземным.

Системы радиосвязи в зависимости от типа обслуживаемых абонентов подразделяются на различные службы. Выделяются системы *фиксированной* и *подвижной* служб, а также *радиовещательной* службы, причем каждая из этих служб может быть как наземной, так и космической (спутниковой). Принято, когда речь идет о наземной службе, не оговаривать в наименовании службы этот факт; специально отмечается лишь принадлежность к спутниковой службе.

Далее, службы делятся в зависимости от типа объектов, для обслуживания которых они предназначены. Так, в зависимости от объектов, на которых размещаются абонентские терминалы, выделяются сухопутная, морская и воздушная подвижные службы. В спутниковой фиксированной службе выделяются *межспутниковая* служба, обеспечивающая связь между спутниками, а также служба *космической эксплуатации*, используемая только в интересах эксплуатации космических кораблей, в частности для управления ими и телеметрирования их характеристик.

Следует отметить, что в регламенте радиосвязи понятие «связь» трактуется широко, так что и радионавигационные, и радиолокационные системы оказываются системами связи. Тем не менее в данной книге таких систем касаться не будем.

Регламент радиосвязи ограничивается классификацией систем, основанной только на их назначении. Для целей стандартизации требований к системам часто используется дополнительная классификация, учитывающая базовые технические принципы их построения. Именно так выделяется класс систем цифровой радиосвязи, которому и посвящена данная книга. К характерной особенности систем этого класса относится то, что информация в них передается в цифровой форме.

Базовой структурой, используемой при построении системы связи, служит одноканальная линия связи, содержащая приемное и передающее оборудование и канал связи. К линии подключаются источник и получатель информации. При этом передающее оборудование может рассматриваться как аппаратура согласования источника с каналом, а приемное — согласования выхода канала с получателем. Общая теория, описывающая передачу цифровой информации по линиям связи, представлена в гл. 2 книги.

В соответствии с теорией информации, согласование источника информации с каналом связи включает процедуры модуляции и кодирования, а выхода канала с получателем — демодуляцию и декодирование. Общая теория и техника кодирования/декодирования изложена в гл. 4, а вопросы демодуляции — в гл. 3. Одному классу методов модуляции — широкополосной манипуляции — посвящен материал гл. 7. В некоторых случаях раздельное рассмотрение процедур кодирования и модуляции оказывается неприемлемым: особенности каждого из используемых при манипуляции сигналов должны учитываться в процессе кодирования для получения эффективных способов передачи информации. Так возникают сигнально-кодовые конструкции, рассмотренные в гл. 5.

В перечисленных выше главах вопросы построения систем связи рассматриваются без учета влияния параллельной работы других систем, использующих тот же связной ресурс. Между тем, радиоканалы практически всегда используются для организации ряда действующих параллельно и одновременно каналов связи. В связи с этим важнейшей проблемой оказывается задача такого построения систем связи, которое сводило бы к допустимому уровню их взаимные влияния при использовании общего связного ресурса. Этой проблеме посвящена гл. 6, в которой изложена общая теория линейного разделения сигналов, проведена классификация методов многостанционного доступа и подробно рассмотрены методы частотно-го, временного и кодового разделения каналов; последнему варианту разделения уделяется значительное место также и в гл. 7.

В двух следующих главах (8 и 9) рассматриваются вопросы борьбы с замираниями, вызванными многолучевым распространением радиоволн. Здесь изучаются модели замираний, относящиеся, в основном, к системам подвижной связи, и методы ослабления их влияния на энергетические показатели радиоканалов.

Заключительная глава учебного пособия (11) призвана показать на важнейших для современной техники связи примерах, как изложенные в книге принципы и методы используются при построении наиболее перспективных систем связи. Здесь рассматриваются системы наземной и спутниковой сотовой связи подвижной службы, а также спутниковой фиксированной службы. По-видимому, этот обзор может представить и самостоятельный интерес.

Несколько особняком смотрится в книге гл. 10, содержащая принципы построения устройств синхронизации систем связи. Это объясняется тем, что по своим теоретическим основам она в значительной степени относится к теории колебаний и теории замкнутых систем автоматического регулирования, чем и отличается от других разделов книги, опирающихся на статистическую радиотехнику. Тем не менее вопросы синхронизации играют столь важную роль в системах связи, что они не могут быть полностью обойдены при изучении структуры таких систем; здесь изложение ограничивается описанием принципов функционирования основных систем синхронизации.

Подводя итог содержанию предлагаемой вниманию читателей книги, отметим, что большая ее часть посвящена системным вопросам, т.е. вопросам, затрагивающим построение системы связи в целом. Лишь гл. 3, 10 и, в определенной степени, гл. 4 рассматривают локальные проблемы, относящиеся к построению отдельных устройств таких систем. При этом и здесь речь идет о важнейших вопросах, определяющих основные характеристики систем связи. Именно эта системная направленность книги и определила ее название.

Глава 1

РАДИОВОЛНЫ, СИГНАЛЫ И ШУМЫ

1.1. Электромагнитные волны. Прием и излучение

1.1.1. Электромагнитные волны

Электромагнитная волна представляет собой векторный процесс, протекающий как во времени, так и в пространстве. Так, полностью характеризующий этот процесс, вектор напряженности электрического поля \vec{E} в общем случае описывается тремя своими проекциями, каждая из которых есть функция от четырех переменных: трех пространственных и времени. При этом действует следующее ограничение: проекция вектора на направление распространения волны всегда равна нулю, так что этот вектор целиком расположен в плоскости, перпендикулярной направлению распространения, именуемой «картинной плоскостью».

В простейшем случае направление вектора \vec{E} электромагнитной волны остается параллельным некоторой прямой для любого момента времени и в любых точках пространства вдоль любого выбранного направления распространения волны. Такая волна называется *линейно поляризованной*. Ясно, что указанная прямая (направление поляризации) должна принадлежать картинной плоскости. В данной плоскости всегда можно выбрать два ортогональных (взаимно перпендикулярных) направления поляризации, что соответствует двум ортогональным линейным поляризациям. При этом любая линейно поляризованная волна, распространяющаяся в данном направлении, может быть представлена как суперпозиция двух заранее зафиксированных ортогонально поляризованных волн. В случае когда волна распространяется вдоль поверхности Земли, часто рассматривают ортогонально поляризованные волны с горизонтальной и вертикальной линейными поляризациями.

Закон изменения вектора \vec{E} в направлении распространения линейно поляризованной волны в фиксированный момент времени представлен на рис. 1.1, причем здесь принято, что направление поляризации лежит в плоскости рисунка.

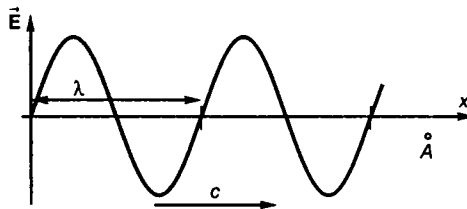


Рис. 1.1. Напряженность поля линейно поляризованной волны

Расстояние между узлами графика, обозначенными засечками на оси абсцисс, есть длина волны λ .

Для того чтобы найти закон изменения \vec{E} во времени, например в точке A , положение которой обозначено на рисунке кружком, необходимо представить себе, что весь график начинает двигаться по стрелке со скоростью света c , а точка A остается неподвижной. При этом, отсчитанная по графику напряженность поля в т. A , будет изменяться по синусоиде, которая и представляет временную зависимость поля в данной точке. Число периодов этой синусоиды в единицу времени (круговая частота волны) определяется соотношением $f = c/\lambda$. Отметим, кстати, что и точка наблюдения A может, вообще говоря, перемещаться. При этом, если проекция ее скорости на направление распространения волны (радиальная скорость) составляет v , частота наблюдаемой в т. A синусоиды составит уже $f_v = \frac{c-v}{\lambda} = f \left(1 - \frac{v}{c}\right)$. Сдвиг частоты колебания для движущегося наблюдателя именуется, как известно, *эффектом Доплера*. Приведенные соотношения показывают, что относительный сдвиг частоты за счет этого явления составляет

$$\delta_d = 1 - \frac{f_v}{f} = \frac{v}{c}. \quad (1.1)$$

Часто в технике используют электромагнитные волны более сложной структуры, чем линейно поляризованные; речь идет о волнах с эллиптической поляризацией. Такие волны формируются в результате суперпозиции двух ортогональных линейно поляризованных волн, распространяющихся в одном и том же направлении. При этом волны должны отличаться также сдвигом местоположений узлов на четверть длины волны. Последнее означает, что если в данный момент времени в данной точке пространства напряженность поля одной из волн проходит через ноль, то для второй в тот же момент в той же точке она проходит через минимум или максимум (в зависимости от направления сдвига).

На рис. 1.2 представлена картина обеих указанных волн в фиксированный момент времени, причем ось x на этом рисунке совпадает с направлением их распространения, y — с направлением поляризации первой волны, а z — второй.

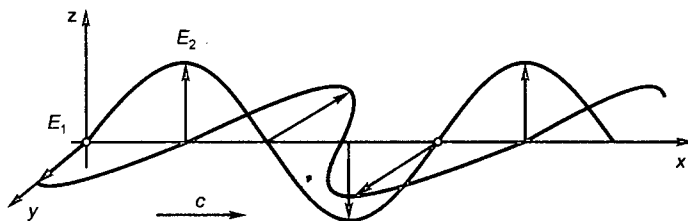


Рис. 1.2. Картина поля волны с эллиптической поляризацией

На рис. 1.3 представлен суммарный вектор напряженности поля первой и второй волн в картинной плоскости в точках оси x с координатой $x_k = k\lambda/4$ в один и тот же момент времени.

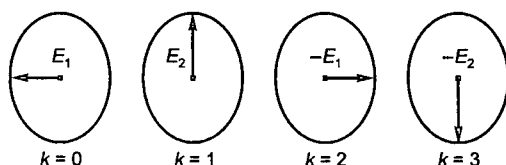


Рис. 1.3. К определению напряженности поля волны с эллиптической поляризацией

Ясно, что при непрерывном перемещении вдоль оси x (момент времени зафиксирован), вектор напряженности суммарного поля вращается, так что его конец описывает представленный на рис. 1.3 эллипс с отношением полуосей $K_3 = E_1/E_2$. Соответственно, суммарная волна называется эллиптически поляризованной, а величина K_3 именуется ее коэффициентом эллиптичности (при определении последней величины в качестве E_1 всегда выбирается меньшая из напряженностей, так что $K_3 \leq 1$).

Такая волна характеризуется также направлением вращения вектора напряженности поля. Принято говорить об эллиптической поляризации правого вращения, если при взгляде на картинную плоскость в направлении распространения волны этот вектор вращается по часовой стрелке; в противном случае вращение считается левым. Легко понять, что направление вращения определяется знаком сдвига фаз волн линейной поляризации, образующих данную эллиптически поляризованную волну. Так, сдвиг на рис. 1.2 дает, как это следует из рис. 1.3, волну с правой поляризацией. Если бы фаза первой волны отставала от второй — поляризация получилась бы левой. Частным случаем эллиптической является круговая поляризация, соответствующая значению коэффициента эллиптичности единица. Круговая поляризация также бывает правого и левого вращений.

1.1.2. Излучение и прием электромагнитных волн

С энергетической точки зрения электромагнитная волна может рассматриваться как процесс переноса энергии от источника излучения в окружающее пространство. Каждую секунду источник через передающую антенну излучает энергию, численно равную его мощности P , и эта энергия уходит в окружающее пространство, подобно расходящимся кругам на воде. Энергию, проходящую при этом через единицу площади картинной плоскости, находящейся на некотором расстоянии R от передающей антенны в единицу времени, называют *вектором Пойнтинга* (размерность Вт/м²). Среднее значение данного вектора по поверхности сферы радиуса R с центром в точке размещения передающей антенны составляет, очевидно, $\bar{W} = \frac{P}{S_{\text{сф}}} = \frac{P}{4\pi R^2}$. Здесь W — модуль вектора Пойнтинга, черта над ним означает ус-

реднение по поверхности сферы, $S_{\text{сф}}$ — площадь этой поверхности. Следует подчеркнуть, что приведенная зависимость имеет место только в идеализированных условиях распространения электромагнитной волны в свободном пространстве. В реальных условиях вид этой зависимости существенно изменяется, о чем будет сказано в следующем пункте.

Поскольку передающие антенны имеют обычно направленные свойства, так что в различные направления излучают сигналы различной интенсивности, в каждом конкретном направлении модуль вектора Пойнтинга не равен среднему значению, а отличается от него в число раз, равное усилению по мощности G передающей антенны в рассматриваемом направлении:

$$W = \bar{W}G = \frac{PG}{4\pi R^2}. \quad (1.2)$$

Выражение (1.2) представляет собой фактически определение коэффициента усиления антенны в любом выбранном направлении, согласно которому, он есть отношение $G = W/\bar{W}$, измеренное в этом направлении.

В соответствии с законом сохранения энергии, чем большая часть мощности источника переносится волной в данном направлении, тем меньшая часть этой мощности излучается

в других направлениях. Поэтому, чем больше коэффициент усиления антенны, тем уже диаграмма ее направленности. Известно следующее ориентировочное соотношение, отображающее описанную зависимость:

$$G = \frac{25\,000}{\alpha_r \alpha_n}, \quad (1.3)$$

где G — коэффициент усиления в направлении максимума диаграммы; α_r и α_n — разворот (в градусах) диаграммы направленности антенны в двух ортогональных плоскостях (по уровню минус 3 дБ).

Перейдем теперь к энергетическим соотношениям при приеме электромагнитных волн. Для ряда антенн, включая и зеркальные, можно исходить из того, что антенна, в идеале без потерь, собирает всю энергию электромагнитной волны, падающую на ее поверхность. Таким образом, мощность на выходе приемной антенны определяется произведением вектора Пойнтинга на площадь проекции антенны на картинную плоскость принимаемой волны:

$$p = WS. \quad (1.4)$$

В выражении (1.4) под S следовало бы понимать геометрическую площадь S_r , если бы вся она одинаково эффективно использовалась антенной. Фактически это в силу ряда причин не так; поэтому в (1.4) должна подставляться так называемая *эффективная площадь антенны*, несколько меньшая S_r :

$$S = \eta S_r, \quad (1.5)$$

где η — коэффициент использования площади, величина по определению меньше единицы (0,5–0,8).

В приведенных соотношениях передающая антенна характеризуется коэффициентом усиления, а приемная — эффективной площадью. Между тем каждая антенна, вообще говоря, может быть использована и как приемная, и как передающая. Поэтому любая антенна может быть эквивалентно охарактеризована любой из указанных двух характеристик. В теории антенн доказана справедливость следующего соотношения между этими характеристиками одной и той же антенны

$$S_1 = \frac{G_1 \lambda^2}{4\pi}. \quad (1.6)$$

Выше отмечалось, что прямая связь эффективной площади антенны с геометрической площадью проекции антенны на картинную плоскость имеет место не всегда. Так, штыревые антенны вообще не имеют площади, а для широко распространенных спиральных антенн сверх того еще существенную роль играет и размер спирали, измеренный вдоль направления распространения. Тем не менее любым антеннам может быть, в соответствии с (1.6), приписана эффективная площадь, после чего по (1.4) можно определить и мощность принимаемого антенной сигнала. При этом если исходным является коэффициент усиления в данном направлении, обозначенном в (1.6) индексом 1, то и эффективная площадь определяется для этого направления и мощность принятого сигнала оценивается при облучении антенны волной с этого направления. Естественно, что для каждой антенны существует такое направление, в котором ее усиление, а следовательно, и эффективная площадь, максимальны. Когда говорят об эффективной площади антенны без указания направления имеют в виду именно максимальную площадь.

Антенны, применяемые в случае излучения и приема радиоволн, рассчитываются на волны определенной поляризации. Так, если в качестве антенны используется штырь, то в идеале он формирует (принимает) только те радиоволны, направление поляризации кото-

рых параллельно этому штырю. С теми же волнами будут согласованы и зеркальные или рупорные антенны, в которых такие штыри используются в качестве облучателей. Очевидно, что антенна, согласованная с данной линейной поляризацией, не должна принимать сигнал ортогональной к ней поляризации.

Могут быть созданы антенны, согласованные с волнами круговой поляризации. Например, можно в качестве такой антенны использовать пару перпендикулярных друг другу штырей, размещенных в картинной плоскости принимаемой волны (рис. 1.4). Сигнал с выхода одного из штырей должен пропускаться через фазовращатель на 90° , после чего суммироваться с выходным сигналом другого штыря.

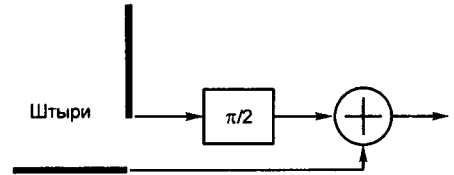


Рис. 1.4. Принцип построения антенны круговой поляризации

Знак фазового сдвига должен быть выбран соответствующим направлению вращения круговой поляризации принимаемого сигнала. При наличии согласования сигналы на входах сумматора окажутся синфазными, так что будет происходить когерентное их суммирование. Если на ту же схему подать волну противоположного вращения, то сигналы на входах сумматора окажутся противофазными и будут в сумматоре полностью компенсироваться. И так, антенна, идеально согласованная с волной круговой поляризации одного направления вращения, совершенно не принимает волну с противоположной круговой поляризацией.

Таким образом, как при линейной, так и при круговой поляризации существует по две ортогональные их разновидности и при том такие, что антенна, идеально согласованная с одной из них, совершенно не реагирует на другую, и наоборот. Легко видеть, что линейная и круговая поляризация между собой ортогональными не являются. Антенна круговой поляризации принимает линейно поляризованную волну и, наоборот, антенна линейной поляризации принимает волну круговой поляризации с небольшим ослаблением (около 3 дБ).

Отметим, что электромагнитная волна с эллиптической поляризацией может быть представлена как сумма двух составляющих с ортогональными круговыми поляризациями. Возможность такого представления поясняется рис. 1.5, где изображены линейно поляризованные составляющие исходной, эллиптически поляризованной волны и две волны с круговой поляризацией, на которые она может быть разложена.

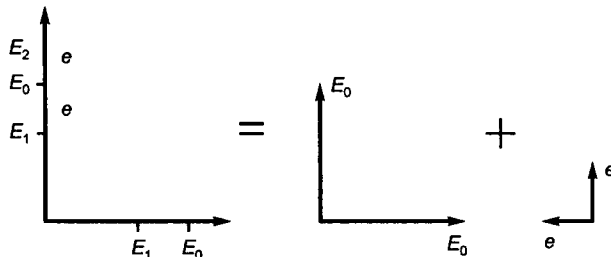


Рис. 1.5. К представлению эллиптически поляризованной волны как суммы волн с круговой поляризацией:

$$E_0 = \frac{E_1 + E_2}{2}; \quad e = \frac{E_2 - E_1}{2}$$

Рис. 1.5 представляет пространственное положение векторов в картинной плоскости волны. Что касается их изменения во времени, то следует помнить, что фаза горизонталь-

ных векторов сдвинута на 90° относительно вертикальных. Однако изменение пространственного положения вектора на противоположное означает одновременно и поворот его фазы во времени на 180° ; поэтому, если первая волна правой части рисунка имеет правую круговую поляризацию, то второй вектор — левую. (Рисунок иллюстрирует равенство волны, представленной в его левой части, сумме двух волн, показанных справа.)

Данные рис. 1.5 показывают, что для волны с коэффициентом эллиптичности K_3 , отношение мощностей составляющих волн основной и ортогональной круговых поляризаций составляет

$$P = \left(\frac{E_0}{e} \right)^2 = \left(\frac{E_1 + E_2}{E_2 - E_1} \right)^2 = \left(\frac{1 + K_3}{1 - K_3} \right)^2. \quad (1.9)$$

Поскольку приемная антенна радиолинии предназначена для приема основной поляризации принимаемой волны, дополнительная (ортогональная) оставляющая ее не принимается. Поэтому передача сигнала с эллиптической поляризацией приводит к потерям мощности сигнала в радиолинии:

$$\eta_3 = (1 + P) / P \quad (1.10)$$

раз, даже при использовании приемной антенны с идеальной круговой поляризацией.

В системах связи часто реализуется возможность повышения эффективности использования полосы рабочих частот за счет одновременной передачи совпадающих по спектру радиоволн на ортогональных поляризациях. При этом идеальное разделение таких волн может быть достигнуто только при идеальной поляризации как в передающей, так и в приемной антеннах. Применительно к круговой поляризации, неидеальность поляризационных свойств антенны определяется коэффициентом эллиптичности волны, создаваемой антенной при работе на передачу. Эту неидеальность можно характеризовать также уровнем кроссполяризации K_n , показывающей во сколько раз (по мощности) антенна, при ее работе на прием, лучше принимает сигнал с той поляризацией, на которую она рассчитана, чем с ортогональной к ней. Вышеизложенное позволяет установить следующую связь между двумя указанными показателями:

$$K_n = P = \left(\frac{1 + K_3}{1 - K_3} \right)^2 \quad (1.11)$$

1.1.3. Энергетический расчет радиолиний

При построении систем радиосвязи важнейшую роль играют энергетические расчеты радиолиний, или, как говорят, анализ их энергетического баланса. Данные предыдущего пункта позволяют произвести такой расчет только в случае, когда можно считать, что радиоволна распространяется в свободном пространстве. Для того чтобы проверить это предположение, в соответствии с результатами теории распространения радиоволн необходимо построить в пространстве эллипсоид вращения с фокусами в точках, где расположены пункты передачи и приема этой радиолинии и с диаметром его сечения посередине между данными пунктами, равным $\sqrt{2\lambda R}$, где λ — длина волны, а R — расстояние между точкам приема и передачи. Если весь объем этого эллипсоида пуст, то можно считать что распространение в данном случае происходит в свободном пространстве. В данном случае вектор Пойнтинга в точке приема определяется выражением (1.2). Если же в объеме эллипсоида присутствуют те или иные предметы, отражающие или поглощающие радиоволны, то энергетические расчеты резко усложняются.

Во-первых, возникает дополнительное затухание волн из-за их поглощения в среде заполняющей эллипсоид связи. Это обстоятельство может быть учтено введением в знаменатель правой части выражения (1.2) некоторого множителя η_1 больше единицы. *Во-вторых*, отражение радиоволн от тех или иных предметов или электрических неоднородностей может привести к тому, что в месте приема будет существовать ряд лучей (многолучевое распространение); возникающая интерференция приведет к изменению коэффициента передачи канала связи. Если при этом происходит перемещение отражающих объектов или объектов между которыми осуществляется связь, то этот коэффициент передачи становится случайным. Для учета этого обстоятельства в выражение (1.2) следует ввести еще один сомножитель η_2 . *В-третьих*, наконец, может оказаться, что прямая трасса между передатчиком и приемником вообще перекрыта практически непрозрачным для радиоволн предметом. При загоризонтном распространении в качестве такого предмета выступает Земля. В этом случае связь возможна только за счет явления дифракции, т.е. огибания волной находящихся на ее пути препятствий. При этом не только резко возрастают потери распространения, но и качественно изменяется их зависимость от дальности связи: вместо квадрата расстояния, фигурирующего в (1.2) появится большее значение степени, так что при загоризонтном распространении

$$W = \frac{PG_1\beta}{4\pi R^{2\nu}}. \quad (1.12)$$

В последнем выражении G_1 — коэффициент усиления (по мощности) передающей антенны в направлении на приемную антенну радиолинии, $\nu \geq 1$; β — величина, размерность которой совпадает с размерностью R^ν . Поэтому, если хотя бы в выражении (1.12) изменить единицы измерения расстояния, то необходимо соответственно изменить и значение β .

С учетом перечисленных обстоятельств в достаточно общем случае вместо (1.12) можно пользоваться соотношением

$$W = \frac{PG_1\beta}{4\pi R^{2\nu}\eta_0}, \quad (1.13)$$

где $\eta = \eta_1 / \eta_2$.

Во всех случаях, когда загоризонтное распространение не используется, в выражении (1.13) полагают $\beta = \nu = 1$. При этом в случае идеального распространения волны в свободном пространстве принимают еще $\eta = 1$.

Используя (1.13), (1.6) и (1.4), найдем выражение для мощности сигнала на выходе приемной антенны:

$$P = \frac{PG_1G_0\lambda^2}{(4\pi)^2 R^{2\nu}\eta}, \quad (1.14)$$

где G_0 — коэффициент усиления (по мощности) приемной антенны радиолинии в направлении на передающую.

Определяемый этим соотношением уровень сигнала на входе приемника приобретает реальный смысл лишь в сопоставлении с уровнем шумов канала связи. Хотя источником шумов (помех) являются различные элементы радиолинии, удобно заменить их все одним эквивалентным источником, действующим на входе приемника (выходе приемной антенны). Мощность такого источника удобно измерять путем ее сопоставления с мощностью дробового шума омического сопротивления, нагретого до некоторой температуры. Температура (в абсолютных градусах Кельвина), до которой следует нагреть такое сопротивление, чтобы при прочих равных условиях мощность его шума стала бы равной мощности шума рассматриваемого источника и именуется температурой шума источника (ниже обо-

значена как $T_{\text{ш}}$). Как известно, мощность шума источника определяется через его шумовую температуру с помощью соотношения

$$p_{\text{ш}} = 1,38 \cdot 10^{-23} T_{\text{ш}} \Delta f, \quad (1.15)$$

где Δf — полоса частот, Гц, в которой определяется мощность шума источника.

Если теперь мощность сигнала (1.14) на выходе приемной антенны радиолинии разделить на мощность шумов, приведенных к той же точке (1.15), то получим основной энергетический параметр радиолинии — отношение сигнал/шум H . Эту величину удобно измерять в дБ. В соответствии с изложенным

$$H [\text{дБ}] = 106 - \eta_0 [\text{дБ}] + 10 \lg b [\text{км}^{2(n-1)}] - 20n \lg R [\text{км}] + \Pi [\text{дБВт}] + U [\text{дБ/К}] - 20 \lg f [\text{ГГц}] - 10 \lg \Delta f [\text{кГц}], \quad (1.16)$$

где $\Pi = PG_1$ — эквивалентная изотропно излучаемая мощность (ЭИИМ) передающего тракта радиолинии; $U = G_0/T_{\text{ш}}$ — приведенная добротность приемной системы радиолинии и f — несущая частота радиолинии. В квадратных скобках указаны размерности величин, которые следует подставлять в (1.16). При этом выбраны размерности более удобные в типовых случаях. По поводу приведенных размерностей отметим еще, что $\Pi [\text{дБВт}] = 10 \lg P [\text{Вт}] + G [\text{дБ}]$ и $U [\text{дБ/К}] = G_0 [\text{дБ}] - 10 \lg T_{\text{ш}} [\text{К}]$.

Нам потребуется еще и выражение для мощности шума на выходе приемной антенны радиолинии, выраженной в дБВт:

$$p_{\text{ш}} [\text{дБВт}] = -198,6 + 10 \lg T_{\text{ш}} [\text{К}] + 10 \lg \Delta f [\text{кГц}]. \quad (1.15a)$$

1.2. Элементы теории вероятностей

1.2.1. Алгебра случайных событий

Современное компактное и логически стройное изложение теории вероятностей требует использования аксиоматики, предложенной А.Н. Колмогоровым. Однако сколько-нибудь строгое ее изложение требует математических знаний, выходящих за рамки уровня, предполагаемого у читателя настоящей книги. В связи с этим, нижеуказанные аксиомы заменены некоторыми, по существу мнемоническими условными соотношениями, определяющими основные понятия теории вероятности. Несмотря на очевидные дефекты такого подхода, он представляется нам целесообразным, поскольку позволяет систематизировать, ясно и кратко напомнить читателю ключевые положения теории вероятностей.

В качестве базовой будем рассматривать схему статистического испытания, т.е. испытания со случайным исходом. Пусть известно множество возможных элементарных исходов испытания (МЭИИ), под которым понимается такое множество, что:

- при каждом испытании осуществляется один и только один из принадлежащих ему исходов;
- все такие исходы равновероятны;
- набор исходов множества достаточен для описания любых событий, которые могут интересовать нас при анализе рассматриваемого испытания.

Представим себе МЭИИ как точки некоторого прямоугольника, который назовем *полем элементарных событий* (ПЭС). Полагаем, что точки равномерно заполняют указанный прямоугольник.

Назовем *событием* некоторое подмножество МЭИИ (некоторую область в ПЭС), считая, что оно происходит тогда и только тогда, когда происходит одно из элементарных событий данного подмножества (области). На рис. 1.6 графически представлено некоторое событие A в ПЭС.

Пусть теперь площадь ПЭС составляет S_{Σ} , а площадь, занимаемая событием A , есть S_A . Тогда по определению вероятностью события A называется величина

$$\mathfrak{Z}\{A\} = S_A / S_{\Sigma}. \quad (1.17)$$

Отметим, что вычисленная таким образом вероятность удовлетворяет условиям

$$0 \leq \mathfrak{Z}\{A\} \leq 1. \quad (1.18).$$

Событие, включающее все точки ПЭС, не входящее в A , называется противоположным A и обозначается \bar{A} . Всякий раз когда выполняется A не выполняется \bar{A} , и наоборот. Очевидно, что

$$\mathfrak{Z}\{\bar{A}\} = 1 - \mathfrak{Z}\{A\}. \quad (1.19)$$

Событие, имеющее нулевую вероятность, называют *невозможным*.

Пару событий в ПЭС называют *несовместными*, если они не пересекаются (события A и C , рис. 1.7) и совместными в противном случае (события A и B). Несовместные события не могут наступить одновременно в каком-либо испытании в противоположность совместным событиям, для которых это возможно. Действительно, это будет происходить, если в данном испытании реализуется элементарное событие, принадлежащее пересечению A и B .

Суммой событий A и B называют новое событие, включающее оба подмножества ПЭС: A и B . Таким образом, событие $D = A + B$ выполняется в данном испытании, если осуществляются или A , или B , или оба события вместе. Область, соответствующая суммарному событию, есть объединение областей A и B .

Произведением событий A и B ($D = AB$) называют событие, соответствующее в ПЭС пересечению множеств A и B . Произведение событий в данном испытании осуществляется только, если одновременно происходят A , и B . *Областью произведения* событий является пересечение областей A и B . Произведение несовместных событий пусто, т.е. является невозможным событием.

Как это следует из данных рис. 1.7, существует следующая связь между вероятностями событий A , B , $A + B$ и AB :

$$\mathfrak{Z}\{A + B\} = \frac{S_A + S_B - S_{AB}}{S_{\Sigma}} = \mathfrak{Z}\{A\} + \mathfrak{Z}\{B\} - \mathfrak{Z}\{AB\}, \quad (1.20)$$

так что вероятность суммы событий равна сумме вероятностей этих событий, только, если события несовместны. (В последнем равенстве площади областей обозначены буквой S с соответствующими индексами.)

При рассмотрении пары событий в теории вероятностей вводится одно из важнейших понятий: понятие условной вероятности (рис. 1.8). Условная вероятность события A при условии B $\mathfrak{Z}\{A/B\}$ представляет собой вероятность того, что если в некотором испытании

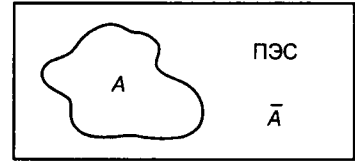


Рис. 1.6. К определению случайного события

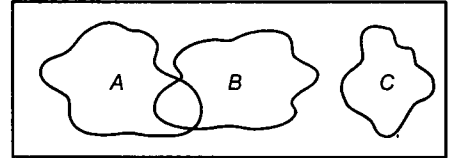


Рис. 1.7. Совместные и несовместные события

выполнилось событие B , то одновременно выполнится и событие A . Очевидно, что при выполнении B событие A выполняется тогда и только тогда, когда выполняется AB . Поэтому рассматриваемая условная вероятность определяется соотношением

$$\mathfrak{P}\{A/B\} = S_{AB} / S_B. \quad (1.21)$$

Это определение позволяет получить важное выражение для вероятности произведения событий:

$$\begin{aligned} \mathfrak{P}\{AB\} &= \frac{S_{AB}}{S_{\Sigma}} = \frac{S_{AB}}{S_A} \frac{S_A}{S_{\Sigma}} = \mathfrak{P}\{A\} \mathfrak{P}\{B/A\} = \\ &= \mathfrak{P}\{B\} \mathfrak{P}\{A/B\}. \end{aligned} \quad (1.22)$$

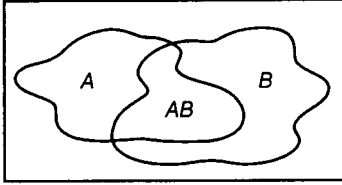


Рис. 1.8. К определению условной вероятности

События A и B называются *независимыми*, если

$\mathfrak{P}\{A/B\} = \mathfrak{P}\{A\}$. Независимость — весьма тонкое свойство событий. Действительно, оно

требует выполнения следующего соотношения: $\frac{S_{AB}}{S_B} = \frac{S_A}{S_{\Sigma}}$, т.е. чтобы площадь AB составля-

ла такую же часть от площади B , как A от всей площади ПЭС. Ясно, что выполнение данного соотношения требует специальной проверки и в общем случае не является заранее очевидным. Важно отметить, что если $\mathfrak{P}\{A/B\} = \mathfrak{P}\{A\}$, то и $\mathfrak{P}\{B/A\} = \mathfrak{P}\{B\}$. Действительно,

из $\frac{S_{AB}}{S_B} = \frac{S_A}{S_{\Sigma}}$ следует $\frac{S_{AB}}{S_A} = \frac{S_B}{S_{\Sigma}}$, что как раз и есть требуемое утверждение. Таким образом,

независимость это взаимное свойство: если событие A не зависит от B , то и B не зависит от A .

В теории вероятности часто полезно использовать набор попарно несовместных событий A_j для $j = 1, 2, \dots, n$, притом таких, что в совокупности они покрывают все ПЭС. Такой набор называют *полным множеством несовместных событий*. Очевидно, что для любого события B имеет место соотношение

$$\mathfrak{P}\{B\} = \sum_{j=1}^n \mathfrak{P}\{BA_j\} = \sum_{j=1}^n \mathfrak{P}\{A_j\} \mathfrak{P}\{B/A_j\}, \quad (1.23)$$

именуемое *формулой полной вероятности*. С ней также тесно связана формула Байеса:

$$\mathfrak{P}\{A_k/B\} = \frac{\mathfrak{P}\{A_k B\}}{\mathfrak{P}\{B\}} = \frac{\mathfrak{P}\{A_k\} \mathfrak{P}\{B/A_k\}}{\sum_{j=1}^n \mathfrak{P}\{A_j\} \mathfrak{P}\{B/A_j\}}. \quad (1.24)$$

1.2.2. Случайные величины и распределения вероятностей

Случайной величиной называется заданная на ПЭС фиксированная функция, принимающая числовые значения. Таким образом, случайная величина в каждом испытании в зависимости от случая принимает то или иное числовое значение.

Полным статистическим описанием случайной величины ξ является ее интегральное распределение вероятностей:

$$P_{\xi}(x) = \mathfrak{P}\{\xi \leq x\} \text{ для всех действительных } x: -\infty < x < \infty. \quad (1.25)$$

Из этого определения следует, что для любого $a > 0$

$$P_{\xi}(x) - P_{\xi}(x - a) = \mathfrak{P}\{x - a < \xi \leq x\}. \quad (1.26)$$

Таким образом, поскольку вероятность величина неотрицательная, интегральное распределение вероятностей всегда не убывает с ростом x , всюду неотрицательно и изменяется от значения 0 для $x \rightarrow -\infty$ до значения 1 для $x \rightarrow \infty$. Общий вид соответствующей функции представлен на рис. 1.9.

На рисунке обращают на себя внимание горизонтальные участки распределения вероятностей. В соответствии с (1.26) они соответствуют интервалам значений случайной величины, которые имеют нулевую вероятность. Так, в интервале от x_2 до x_3 случайная величина значений никогда не принимает.

Заслуживает анализа также поведение распределения в точке x_1 , где эта функция терпит разрыв первого рода. В согласии с (1.26) это поведение отображает тот факт, что рассматриваемая случайная величина принимает значение x_1 с конечной вероятностью b . Если же в точке x распределение непрерывно, то это означает, что вероятность того, что случайная величина примет значение в точности равно x , стремится к нулю. Соответственно различают дискретные случайные величины, которые могут принимать лишь отдельные значения, так что для них $P_\xi(x)$ состоит лишь из ступеней (скачков, соединяемых горизонтальными участками) и непрерывно распределенные величины, для которых $P_\xi(x)$ непрерывная функция. Существуют еще и смешанные случайные величины; на рис. 1.9 изображено распределение вероятностей именно такой величины.

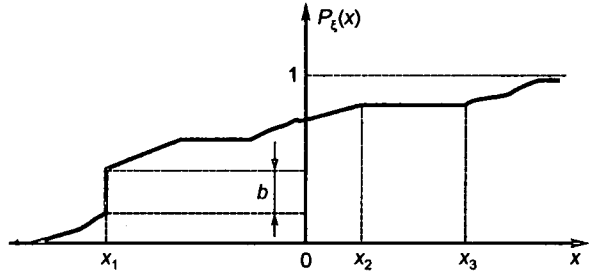


Рис. 1.9. Интегральное распределение вероятностей случайной величины

Часто вместо $P_\xi(x)$ используют плотность распределения вероятностей

$$p_\xi(x) = \frac{d}{dx} P_\xi(x). \quad (1.27)$$

В случаях наличия дискретной составляющей в $P_\xi(x)$, как отмечалось выше, имеются точки разрыва непрерывности, так что производная в обычном смысле не существует. Чтобы обойти эту трудность, в теории вероятностей договариваются об использовании обобщенной δ -функции Дирака. Последнюю можно представлять как бесконечно короткий импульс, но столь большой амплитуды, что его площадь равна единице. Основное свойство этой функции задается соотношением

$$\int_{-a}^a \delta(x) f(x) dx = f(0), \quad (1.28)$$

где $f(x)$ — любая функция.

Ясно, что через $\delta(x)$ можно выразить производную от функций, имеющих разрывы первого рода. В частности, производная от функции, представленной на рис. 1.9, в точке x_1 есть $p_\xi(x_1) = b\delta(x - x_1)$. Как это следует из определения (1.27), плотность распределения вероятностей всюду неотрицательна, и интеграл от нее по всей области определения равен единице.

Перейдем теперь к рассмотрению пары случайных величин ξ и ζ , заданных на одном и том же ПЭС. Ее можно рассматривать как случайный вектор на плоскости. Полное статистическое описание такого вектора дается двумерным распределением вероятностей, которое определяется соотношением

$$P_{\xi, \zeta}(x, y) = \mathfrak{F}(\xi \leq x, \zeta \leq y). \quad (1.29)$$

Эта величина характеризует вероятность того, что точка с координатами ξ и ζ лежит внутри угла, представленного на рис. 1.10, а.

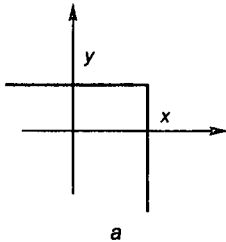
Вычислим теперь вероятность того, что эта точка принадлежит прямоугольнику с координатами вершин (a, d) , (a, c) , (c, b) и (b, d) (рис. 1.10, б)

$$\mathfrak{P}\{a < \xi \leq b, c < \zeta \leq d\} = P_{\xi, \zeta}(b, d) - P_{\xi, \zeta}(a, d) - P_{\xi, \zeta}(b, c) + P_{\xi, \zeta}(a, c). \quad (1.30)$$

Из этого выражения следует, что вероятность q попадания двумерной случайной величины в достаточно малый прямоугольник со сторонами dx , dy с центром в точке с координатами x , y определяется соотношением

$$q = p_{\xi, \zeta}(x, y), \quad (1.31)$$

где обозначено $p_{\xi, \zeta}(x, y) = \frac{\partial^2}{\partial x \partial y} P_{\xi, \zeta}(x, y)$; эта величина именуется двумерной плотностью распределения вероятностей.



а

Вероятность (1.30) может быть выражена через эту величину в виде

$$\mathfrak{P}\{a < \xi \leq b, c < \zeta \leq d\} = \int_a^b dx \int_c^d p_{\xi, \zeta}(x, y) dy. \quad (1.32)$$

Часто используются представления двумерных распределений через условные распределения вероятностей:

$$p_{\xi, \zeta}(x, y) = p_{\xi}(x) p_{\zeta|\xi}(y). \quad (1.33)$$

Для независимых случайных величин ξ и ζ :

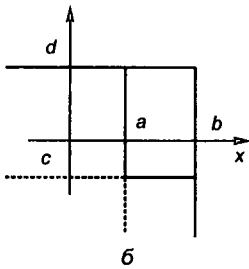
$$p_{\xi, \zeta}(x, y) = p_{\xi}(x) p_{\zeta}(y); \quad p_{\xi, \zeta}(x, y) = p_{\xi}(x) p_{\zeta}(y). \quad (1.34)$$

Вместо полных статистических характеристик случайных величин широко используют их числовые параметры, нахождение которых основано на процедуре взятия математического ожидания. Математическое ожидание случайной величины ξ по определению составляет

$$M\xi = \bar{\xi} = \int_{-\infty}^{\infty} xp_{\xi}(x) dx \quad (1.35)$$

(символом M или чертой сверху обозначается математическое ожидание стоящей за/под ним величины). Происхождение этого выражения следующее: если выбрать достаточно малый шаг Δ и разбить всю область возможных значений

Рис. 1.10. К определению двумерного интегрального распределения вероятностей



б

случайной величины на интервалы с таким шагом, то вероятность того, что случайная величина окажется в некотором из них с центром в x_j ($j = \dots -1, 0, 1, \dots$) составит $q_i = \Delta p_{\xi}(x_i)$. Если испытания по реализации рассматриваемой случайной величины произвести достаточно большое число раз N , то в j -й интервал случайная величина попадет приблизительно $n_i = Nq_i = N\Delta p_{\xi}(x_i)$ раз. Поэтому, если сложить теперь все полученные в этих испытаниях значения случайной величины и разделить эту сумму на число испытаний, т.е. определить среднее значение случайной величины по всем испытаниям, то получим число, приближающееся к даваемому определением (1.35). Приближение при этом будет тем лучшим, чем меньше шаг Δ и больше число испытаний N .

Отметим два очевидных свойства математического ожидания: для неслучайного числа a имеем $Ma = a$ и $M a\xi = a\xi$. Далее, если рассматривается случайная величина, получаемая

как произведение двух независимых случайных величин $\eta = \xi\zeta$, то $M\eta = (M\xi)(M\zeta)$. Важными являются два других свойства, которые требуют уже специального доказательства. Во-первых, если рассматриваются две случайные величины ξ и ζ произвольным образом связанные друг с другом, то

$$M(\xi + \zeta) = M\xi + M\zeta. \quad (1.36)$$

Далее, если рассматривается любая фиксированная (неслучайная) функция от случайной величины $f(\xi)$, то

$$M[f(\xi)] = \int_{-\infty}^{\infty} f(x)p_{\xi}(x)dx. \quad (1.37)$$

Смысл последнего соотношения заключается в следующем: $\zeta = f(\xi)$ есть уже новая случайная величина, отличная от ξ , так что вычисление ее математического ожидания по формуле (1.35) требует предварительного определения ее распределения вероятностей. Последнее, хотя и определяется однозначно по распределению исходной величины ξ , но отнюдь не совпадает с ним. Так вот, использование выражения (1.37) позволяет избежать пересчета распределений вероятностей.

Отметим теперь, что случайная величина с $M\xi = 0$ называется *центрированной*. Для каждой случайной величины ξ может быть построена соответствующая ей центрированная величина $\xi_0 = \xi - \bar{\xi}$.

Важной числовой характеристикой случайной величины является ее дисперсия, по определению равная $D\xi = \sigma^2 = M(\xi - \bar{\xi})^2$ (σ^2 ее общепринятое обозначение, причем σ именуется среднеквадратичным значением ξ). В соответствии с указанными свойствами математического ожидания имеем

$$D\xi = M(\xi^2) - (M\xi)^2 = \int_{-\infty}^{\infty} x^2 p_{\xi}(x)dx - (\bar{\xi})^2. \quad (1.38)$$

Среди различных случайных величин исключительную роль играет нормальная (гауссова) случайная величина, плотность распределения вероятности которой имеет вид

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right], \quad (1.39)$$

где a и σ^2 — математическое ожидание и дисперсия случайной величины. Эта особая роль определяется тем фактом, что распределение суммы произвольных случайных величин приближается к нормальному по мере роста их числа, если только неограниченно возрастает отношение среднеквадратичного значения суммы к среднеквадратичному значению каждого из ее членов (центральная предельная теорема теории вероятностей).

В радиотехнике часто используется специальная функция

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp(-x^2/2)dx, \quad (1.40)$$

именуемая *функцией ошибок*. Поведение этой функции определяется данными табл. 1.1.

Таблица 1.1. Функция ошибок

$\Phi(x)$	0,5	0,3	0,2	0,1	0,03	0,01	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}
x	0	0,525	0,843	1,21	1,88	2,33	3,09	3,72	4,27	4,75	5,25

Для больших x известно следующее асимптотическое выражение для функции ошибок:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}x} \left(1 - \frac{1}{x^2} \right) \exp(-x^2/2). \quad (1.41)$$

Интегральное распределение нормальной случайной величины, соответствующее (1.39), определяется через функцию ошибок соотношением

$$P(x) = 1 - \Phi\left(\frac{x-a}{\sigma}\right). \quad (1.42)$$

Из случайных величин, получаемых в результате преобразования двух других случайных величин, рассмотрим μ — длину случайного вектора на плоскости, декартовы координаты которого независимые нормальные случайные величины. Примем, что дисперсии координат одинаковы и равны σ^2 , а их математические ожидания составляют a и b , причем $\mu_0 = \sqrt{a^2 + b^2}$. Плотность распределения величины μ :

$$p_\mu(x) = \frac{x}{\sigma^2} I_0\left(\frac{\mu_0 x}{\sigma^2}\right) \exp\left(-\frac{x^2 + \mu_0^2}{2\sigma^2}\right), \quad (1.43)$$

где $I_0(y)$ — функция Бесселя от мнимого аргумента нулевого порядка.

Это распределение называется *распределением Райса*. Частный случай этого распределения, соответствующий $a = b = 0$, носит название *распределения Релея* и имеет вид

$$p_\mu(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right). \quad (1.43a)$$

1.2.3. Случайные процессы

Перейдем теперь к рассмотрению нового статистического объекта — случайных процессов. Реализацией интересующих нас случайных процессов являются функции времени $\xi(t)$, причем появление той или иной реализации зависит от случая. Значения случайного процесса можно рассматривать как множество случайных величин $\xi(t)$ (при каждом значении t своя случайная величина).

Полное статистическое описание случайного процесса должно позволять построить совместные распределения вероятностей любого конечного числа n отсчетов процесса $\xi(t_k) = \xi_k$, взятых в некоторые (любые) моменты времени t_k ($k = 1, 2, \dots, n$). Если такое совместное распределение зависит только от взаимного расположения точек t_k , но не зависит от общего сдвига всех точек по оси времени, то процесс называется *стационарным*. Далее, если математическое ожидание любых функций от ξ_k , вычисленное по указанному выше совместному распределению, совпадает со средним значением этих функций по каждой реализации этого процесса, то последний именуется *эргодическим*. (Для эргодических процессов каждая реализация полностью представляет процесс, поскольку позволяет определить все его статистические характеристики.)

Получение полного статистического описания случайного процесса задача весьма сложная и трудоемкая, в связи с чем мы ограничимся только теми сведениями о процессе, которые можно получить, опираясь на его двумерное распределение, т.е. совместное распределение пар его отсчетов: $\xi(t)$ и $\xi(t - \tau)$ для любых τ .

Прежде всего, даже по одномерному распределению могут быть определены математическое ожидание и дисперсия $\xi(t)$. Для стационарных процессов эти величины не зависят от t ; для нестационарных, но эргодических их имеет смысл усреднить еще по времени:

$$a = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \overline{\xi(t)} dt \text{ и } \sigma^2 = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \overline{[\xi(t) - \overline{\xi(t)}]^2} dt. \quad (1.44)$$

Процесс с $\overline{\xi(t)} = 0$ называют центрированным. Если не оговорено противное, будут рассматриваться центрированные процессы. Важной характеристикой процесса является его функция корреляции

$$R(t, \tau) = M[\xi(t)\xi(t - \tau)]. \quad (1.45)$$

Для стационарных процессов эта функция не зависит от t , а для эргодических, но не стационарных, имеет смысл усреднить ее по времени:

$$R(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T R(t, \tau) dt. \quad (1.46)$$

Функция корреляции обладает следующими свойствами. Во-первых, как это следует из определения (1.35), она является четной функцией τ :

$$R(\tau) = R(-\tau). \quad (1.47)$$

Во-вторых,

$$R(0) = \sigma^2 \geq |R(\tau)|. \quad (1.48)$$

Как правило, с ростом τ статистическая связь между $\xi(t)$ и $\xi(t - \tau)$ убывает, так что при достаточно большом (по модулю) τ эти случайные величины могут считаться независимыми. Для таких τ имеем

$$R(\tau) = \overline{\xi(t)\xi(t - \tau)} = \overline{\xi(t)} \overline{\xi(t - \tau)} = 0. \quad (1.49)$$

Поэтому значения τ , при которых значения функции корреляции существенно убывают, по отношению к значению при $\tau = 0$ ($R(0) = \sigma^2$), указывает на интервал времени, в течение которого реализация процесса успевает существенно измениться. Эти соображения качественного характера показывают, что функция корреляции определяет скорость изменения реализаций случайного процесса, а следовательно, и параметры его спектра. Соответствующее точное утверждение заключается в том, что спектр по Фурье функции корреляции есть энергетический спектр случайного процесса:

$$N_0(f) = 2 \int_{-\infty}^{\infty} R(\tau) e^{-2\pi f \tau} d\tau = 4 \int_0^{\infty} R(\tau) \cos(2\pi f \tau) d\tau. \quad (1.50)$$

Таким образом, $N_0(f)$ принимает только действительные значения. В соответствии с обратным преобразованием Фурье из (1.50) следует

$$R(\tau) = \int_0^{\infty} N_0(f) \cos(2\pi f \tau) df. \quad (1.51)$$

Физический смысл энергетического спектра определяется следующим: мощность составляющих случайного процесса с частотами, лежащими в полосе от f_1 до f_2 , определяется соотношением

$$\sigma_{1,2}^2 = \int_{f_1}^{f_2} N_0(f) df. \quad (1.52)$$

В соответствии с указанным физическим содержанием энергетический спектр ни при каком значении f не может иметь отрицательных значений. Отсюда следует весьма тонкое свойство функции корреляции: ею могут быть только такие функции, преобразование

Фурье которых всюду неотрицательно. Так, функция корреляции не может иметь форму прямоугольного импульса, поскольку его преобразование Фурье — функция вида $\sin x/x$ принимает на определенных интервалах аргумента отрицательные значения. Наоборот, треугольный импульс может быть функцией корреляции, поскольку его спектр имеет вид $(\sin x/x)^2$ и поэтому всюду неотрицателен.

В теории связи часто используют некоторый идеализированный случайный процесс $\alpha(t)$, именуемый *белым гауссовским шумом* (БГШ), для которого энергетический спектр постоянен на всех частотах, а математическое ожидание равно нулю. В соответствии с (1.50) для БГШ

$$R(\tau) = \frac{N_0}{2} \delta(\tau), \quad (1.53)$$

где $\delta(x)$ — δ -функция Дирака. Часто встречаются также случайные величины и процессы, образованные в результате тех или иных преобразований из БГШ. Так, можно рассмотреть

случайную величину $\xi = \int_0^T \alpha(t)b(t)dt$, где $b(t)$ — некоторая фиксированная неслучайная

функция времени. Величина ξ центрирована (из свойства математического ожидания суммы следует, что математическое ожидание интеграла равно интегралу от математического ожидания):

$$M\xi = M \int_0^T \alpha(t)b(t)dt = \int_0^T M[\alpha(t)]b(t)dt = 0.$$

Ее дисперсия составляет

$$\begin{aligned} \sigma^2 &= M \left[\int_0^T \alpha(t)b(t)dt \right]^2 = M \left\{ \left[\int_0^T \alpha(x)b(x)dx \right] \left[\int_0^T \alpha(y)b(y)dy \right] \right\} = \int_0^T \int_0^T \overline{\alpha(x)\alpha(y)} b(x)b(y)dy dx = \\ &= \int_0^T \int_0^T R(x-y)b(x)b(y)dx dy = \frac{N_0}{2} \int_0^T \int_0^T \delta(x-y)b(x)b(y)dx dy = \frac{N_0}{2} \int_0^T b^2(x)dx. \end{aligned} \quad (1.54)$$

Все приведенные выше данные о случайных процессах базировались только на использовании их двумерных распределений вероятностей. Между тем нам потребуются и некоторые данные, касающиеся вида реализаций процесса на всем интервале его определения. Можно задаться вопросом о вероятности того, что реализация процесса примет ту или иную форму, например, будет описываться некоторой заранее заданной функцией $\alpha(t)$. Следует сразу же отметить, что в такой постановке вопрос имеет, как правило, тривиальный ответ — эта вероятность равна нулю; просто число возможных реализаций процесса столь велико, что вероятные появления любой конкретной из них равна нулю. Для того чтобы получить содержательные утверждения, целесообразно несколько изменить рассматриваемый вопрос. Попытаемся выяснить, во сколько раз вероятность появления некоторой конкретной реализации $\alpha(t)$ больше (меньше), чем некоторой другой реализации $\alpha_0(t)$, принятой за базовую. Такой вопрос является уже вполне содержательным. Ответ на него дает, так называемый *функционал плотности вероятности* (ФПВ) случайного процесса $W[\alpha(t)]$. Эта величина называется функционалом, поскольку представляет собой функцию от функции. Она определяется однозначно, если зафиксирована базовая реализация. Если же эта реализация не фиксируется, то функционал определяется с точностью до постоянного множителя. Каждый процесс имеет свой ФПВ. Ниже приведен вид такого функционала для БГШ:

$$W[\alpha(t)] = c \exp \left[-\frac{1}{N_0} \int_0^T \alpha^2(t)dt \right], \quad (1.55)$$

где c — константа, зависящая от выбора базовой реализации. Из этого соотношения следует, что, чем меньше энергия реализации, тем больше вероятность ее появления. Наиболее вероятной является реализация всюду равная нулю. Может возникнуть вопрос: почему, если это так, мы никогда не наблюдаем на выходе линейной части радиоприемника, например, чисто нулевой реализации шумов. Дело в том, что нулевая реализация, хотя и наиболее вероятна, но единственна, в то время как ненулевых реализаций бесконечное число, так что их суммарная вероятность равна единице, что соответствует достоверному событию.

1.3. Видео- и радиосигналы. Цифровая модуляция

1.3.1. Видео- и радиосигналы

Сообщения, транслируемые по каналам связи, являются случайными (определяются внешними источниками информации), в связи с чем адекватное общее описание сигналов, циркулирующих в этих каналах, должно опираться на теорию случайных процессов. В частности, спектральный состав сигналов описывается их энергетическими спектрами.

С теоретической точки зрения полоса частот, занимаемая связным сигналом, не может быть строго ограниченной. Дело в том, что сигналы со строго ограниченным спектром не могут быть тождественно равны нулю ни на каком конечном временном интервале. Кроме того, поскольку они являются аналитическими функциями (имеющими производные любых порядков), их можно в точности предсказать на любое время вперед по значениям на некотором конечном отрезке времени. Ясно, что реальные сигналы, всегда начинающиеся в определенный момент времени и допускающие непредсказуемые изменения по воле пользователя, не удовлетворяют этим требованиям.

Тем не менее для реальных сигналов всегда можно так выбрать конечный интервал частот, чтобы их спектральные составляющие, лежащие вне этого интервала, имели бы столь малый уровень, что применительно к той или иной конкретной задаче ими можно было бы пренебречь. В связи с этим мы часто будем далее считать спектр рассматриваемых сигналов ограниченным. Подчеркнем лишь, что вопрос о полосе частот, занимаемых этими сигналами, т.е. об уровнях тех их спектральных компонент, которыми можно уже пренебречь, должен всегда специально анализироваться исходя из конкретной рассматриваемой задачи.

Итак, обозначим полосу частот, занятых сигналом, через Δf , а среднюю частоту его спектра через f_0 . Видеосигналами будем называть сигналы, у которых Δf соизмеримо с f_0 . Если же $f_0 \gg \Delta f$, то будем говорить о радиосигналах (рис. 1.11). Подчеркнем, что эта классификация вовсе не зависит от абсолютных значений полосы и средней частоты, а определяется только их соотношением.

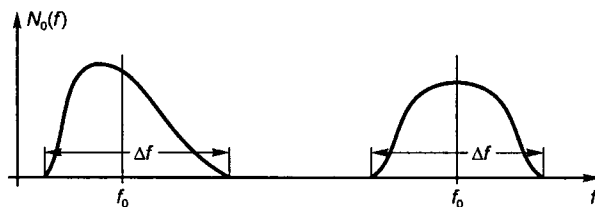


Рис. 1.11. К определению радио- и видеосигналов

В зависимости от вида областей определения и значений функций, описывающих сигналы, последние делятся на аналоговые, дискретные и цифровые. Аналоговые сигналы определены на всей оси времени и принимают любые значения из некоторого интервала уровней (например, от $-a$ до a). Дискретные сигналы определены только в моменты времени $t_k = kT + t_0$, где T — длительность такта; t_0 — начальный временной сдвиг, именуемый также сдвигом по такту; k — любое целое число, положительное, отрицательное или нуль. Моменты времени t_k именуются *тактовыми точками*. Область значений дискретных сигналов такая же, как и у аналоговых. Для цифровых сигналов — область определения такая же, как и у дискретных, но область значений содержит лишь конечное число точек. При этом, если таких точек n , то говорят об n -ичных цифровых сигналах. Если $n = 2^k$, то сигналы именуется k -кратными. На рис. 1.12 показаны аналоговые, дискретные и цифровые сигналы соответственно.

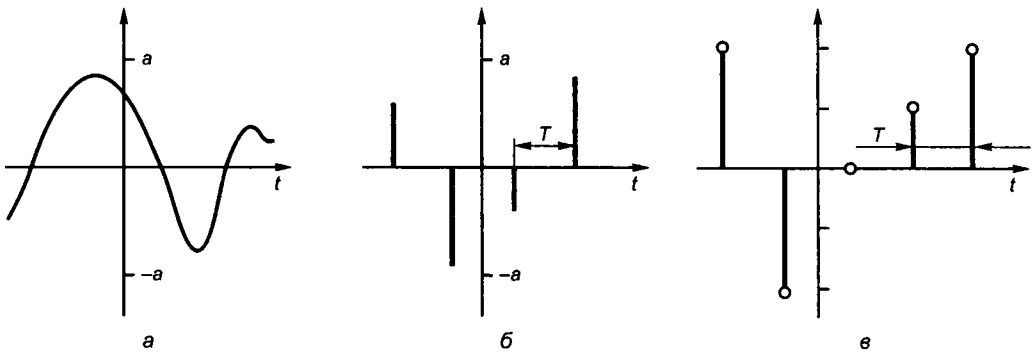


Рис. 1.12. Аналоговые (а), дискретные (б) и цифровые (в) сигналы

Сигналы, на входах (выходах) систем связи, отображающие транслируемую этой системой информацию, будем называть *сообщениями*. Источники сообщений могут формировать их в любой форме: аналоговой дискретной или цифровой. Однако поскольку настоящая книга посвящена цифровой передаче информации, здесь всегда будет предполагаться, что перед передачей сообщения преобразуются в цифровую форму, так что в качестве передаваемых сообщений будут рассматриваться цифровые сигналы. В результате модуляции (применительно к цифровым сообщениям этот процесс именуется манипуляцией, или телеграфией) сообщения преобразуются в аналоговые радиосигналы, которые собственно и передаются по каналам связи. На приемной стороне радиосигналы демодулируются, в результате чего и формируется принятое цифровое сообщение. Аналоговые радиосигналы могут быть записаны в виде

$$u(t) = A(t) \cos[\omega t + \varphi(t)], \quad (1.56)$$

где $A(t)$ — огибающая сигнала; $\varphi(t)$ — закон изменения его фазы и ω — опорная частота, принадлежащая полосе частот, занятых сигналом. Процесс манипуляции сводится к изменению одного или нескольких параметров радиосигнала в соответствии с транслируемым сообщением. С учетом манипулируемого параметра методы манипуляции носят названия амплитудная, частотная, фазовая, частотно-фазовая и т.п. Подробнее основные методы манипуляции будут описаны ниже. При представлении радиосигнала в виде (1.56), входящие в это выражение параметры не определяются единственным образом. Во-первых, про-

извольным является выбор опорной частоты. Фактически однозначно определяется лишь полная фаза $\theta(t) = \omega t + \varphi(t)$, так что изменение ω приводит к соответствующему изменению $\varphi(t)$, сводящемуся к добавлению к этой фазе линейно нарастающей (убывающей) составляющей. Можно, однако, так выбрать опорную частоту, чтобы $\varphi(t)$ не содержало такой составляющей, т.е. чтобы среднее значение производной от $\varphi(t)$ равнялось нулю. Для сигналов с симметричным энергетическим спектром такой частотой является средняя частота спектра ω_0 . Далее в связи с (1.56) следует уточнить понятие огибающей. При строгом подходе естественно потребовать, чтобы $A(t)$ было бы всегда неотрицательным. Иногда удобно допускать и отрицательные значения $A(t)$, означающие фактически дополнительный поворот фазы $\varphi(t)$ на 180° . Чтобы различать эти два подхода, не вводя новые термины, в общем случае будем называть $A(t)$ амплитудой, сохраняя термин огибающая для случая, когда эта величина вычислена так, что ее значения неотрицательны.

Для облегчения аппаратурной реализации процедур обработки радиосигналов эти сигналы, как правило, заменяют соответствующими им видеосигналами. Соответствие понимается здесь в том смысле, что по этим видеосигналам могут быть найдены функции $A(t)$ и $\varphi(t)$, однозначно определяющие радиосигнал (при известной опорной частоте ω). Для пояснения одного из возможных вариантов построения видеозэквивалента радиосигнала рассмотрим как изменяются $A(t)$ и $\varphi(t)$ при переносе радиосигнала на промежуточную частоту $f_{\text{пр}}$. Пока эта частота остается больше $\Delta f/2$, спектральные зоны, лежащие в области положительных и отрицательных частот, хотя и сближаются, но не пересекаются, так что искажения вида спектра сигнала не происходит и $A(t)$ и $\varphi(t)$ сохраняются неизменными. Искажения этих функций наступают тогда, когда $f_{\text{пр}}$ оказывается меньше $\Delta f/2$, так что указанные зоны начинают пересекаться (рис. 1.13)

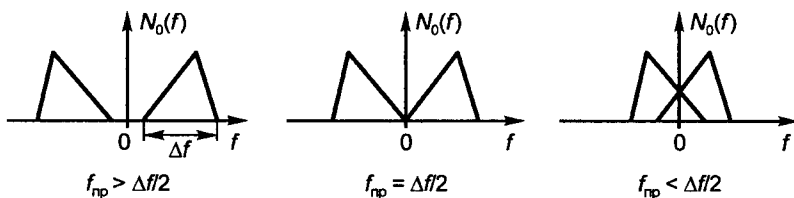


Рис. 1.13. К определению понятия видеозэквивалентов радиосигнала

Случай $f_{\text{пр}} = \Delta f/2$ является в этом смысле предельным. Получаемый сигнал промежуточной частоты будем условно именовать *предельным видеозэквивалентом* (ПВЭ) данного радиосигнала. Так, ПВЭ представляет собой видеосигнал, занимающий полосу частот от 0 до Δf . В соответствии с (1.56) он может быть записан аналитически в виде

$$v(t) = A(t) \cos[\pi \Delta f t + \varphi(t)]. \quad (1.57)$$

Подчеркнем, что спектр исходного радиосигнала перед формированием ПВЭ должен быть достаточно эффективно ограничен полосой Δf . Такое ограничение требует использования перед преобразователем частоты аналогового полосового фильтра, в достаточной степени подавляющего спектральные компоненты сигнала, выходящие за полосу Δf . Поскольку создание таких достаточно узкополосных фильтров может оказаться сложным, часто идут на то, чтобы увеличивать полосу Δf сверх истинной полосы сигнала в соответствии с возможностями фильтра.

Другая возможность получения видеосигналов, соответствующих данному радиосигналу, следует из преобразования представления (1.56):

$$\begin{aligned}
 u(t) &= A(t) \cos[\omega t + \varphi(t)] = A(t) \cos[\varphi(t)] \cos \omega t - A(t) \sin[\varphi(t)] \sin \omega t = \\
 &= Q(t) \cos \omega t + I(t) \sin \omega t,
 \end{aligned} \tag{1.58}$$

где $Q(t) = A(t) \sin \varphi(t)$ и $I(t) = -A(t) \cos \varphi(t)$ — видеосигналы, именуемые *квадратурными координатами* (КК) радиосигнала.

В соответствии с (1.58) произвольный радиосигнал может быть представлен как сумма двух амплитудно модулированных колебаний, несущие которых ($\sin \omega t$ и $\cos \omega t$) сдвинуты по фазе на 90° , а законами амплитудной модуляции служат квадратурные координаты $I(t)$ и $Q(t)$ соответственно. (Отметим, что указанные координаты могут быть как положительными, так и отрицательными.) Представление радиосигнала (1.58) именуется *квадратурным*. Легко видеть, что если $\omega \approx \omega_0$, то КК радиосигнала являются видеосигналами. Эти координаты могут быть определены путем преобразования радиосигнала на нулевую промежуточную частоту. Действительно, если умножить (1.58) на $\cos \omega t$, получим

$$y(t) = u(t) \cos \omega t = Q(t) \cos^2 \omega t + I(t) \sin \omega t \cos \omega t = \frac{1}{2} Q(t) + \frac{1}{2} Q(t) \cos 2\omega t + \frac{1}{2} I(t) \sin 2\omega t.$$

С учетом полос частот, занимаемых КК, можно утверждать, что члены частоты 2ω , входящие в это выражение, могут быть легко отфильтрованы. Для этого должен быть использован фильтр нижних частот, пропускающий без искажения спектральные компоненты с частотами в диапазоне от $-\Delta f/2$ до $\Delta f/2$ и эффективно подавляющий частоты вблизи частоты 2ω . Если процедуру такой фильтрации обозначить оператором \mathbf{H} то получим

$$Q(t) = \mathbf{H} u(t) \cos \omega t. \tag{1.59}$$

Поступая аналогично, найдем

$$I(t) = \mathbf{H} u(t) \sin \omega t. \tag{1.60}$$

Из этих соотношений следует, что КК занимают полосу частот от $-\Delta f/2$ до $\Delta f/2$.

1.3.2. Цифровые представления видеосигналов

Остановимся на вопросах преобразования аналоговых видеосигналов в цифровую форму (АЦП). Способы реализации АЦП можно разделить на две группы: специализированных и универсальных преобразований. *Специализированные* преобразования существенно используют конкретные свойства данного источника. Специализированными являются, например, вокодеры, используемые для преобразования в цифровую форму речевых сигналов, или преобразователи телевизионных изображений, построенные с использованием стандарта MPEG. При специализированных преобразованиях за счет использования особенностей как самих сообщений, так и их восприятия пользователем обеспечивается высокая эффективность преобразования. Зато такие преобразования оказываются неудобными, когда канал связи предназначается для трансляции сообщений различных источников. Так, для телефонных каналов (каналов тональной частоты — ТЧ), кроме передачи речевых сообщений, должна предусматриваться еще и трансляция факсимильных и цифровых сообщений, соответствующим образом преобразованных для согласования с этими каналами. Между тем вокодер, весьма эффективный при АЦП речевых сигналов, практически не пригоден для обработки факсимильных и цифровых сообщений.

По этой причине широко применяются также универсальные способы АЦП, для которых класс обрабатываемых сообщений ограничивается лишь заданием ширины их спектра и динамическим диапазоном. Ниже остановимся на одном из них, которое традиционно именуется ИКМ-преобразованием.

ИКМ-преобразование основано на непосредственном использовании теоремы Котельникова, в соответствии с которой видеосигнал $x(t)$ с ограниченным спектром, занимающий полосу частот от $-F/2$ до $F/2$, однозначно определяется своими отсчетами, т.е. значениями $a_k = x(t_k)$, взятыми в точках $t_k = t_0 + kq$ ($k = \dots, -1, 0, 1, 2, \dots$ — номер отсчета; t_0 — любое число, именуемое начальным сдвигом отсчетов), если только $q \leq 1/F$. Именно, имеет место равенство

$$x(t) = \sum_{k=-\infty}^{\infty} a_k \frac{\sin[\pi(t-t_k)/q]}{\pi(t-t_k)/q}. \quad (1.61)$$

Соотношение (1.61) именуется *рядом Котельникова*.

Сформулированное соотношение означает, что аналоговый видеосигнал $x(t)$ в точности определяется дискретным сигналом $\{a_k\}$, представляющим собой множество его отсчетов. Поэтому $\{a_k\}$ может рассматриваться как дискретное представление аналогового видеосигнала. Процедура получения дискретного представления носит название *дискретизации*.

Точное равенство (1.61) имеет место только при строгом ограничении ширины спектра аналогового сигнала. Однако, как отмечалось выше, это условие на практике не имеет места. В связи с этим реально (1.61) выполняется лишь приближенно. Если $x(t)$ случайный процесс с энергетическим спектром $S(f)$, то дисперсия ошибки дискретизации, как показано в литературе, составляет

$$\sigma_d^2 = 2 \int_{F/2}^{\infty} S(f) df. \quad (1.62)$$

Дискретное представление может быть заменено цифровым, если значения отсчетов a_k так или иначе округлить. Процедура такого округления именуется *квантованием по уровню*, или просто *квантованием*. В общем виде она описывается следующим образом: выбирается некоторое число порогов квантования A_i ($i = 0, 1, \dots, N+1$), причем так, чтобы каждый следующий порог превосходил предыдущий. При этом всегда $A_0 = -\infty$, а $A_{N+1} = \infty$. Интервалы между соседними порогами квантования именуется *интервалами квантования*. На каждом интервале выбирается одно значение, называемое *уровнем квантования* b_i ($i = 1, 2, \dots, N$). Каждый раз когда отсчет a_k попадает в i -й интервал квантования, его истинное значение округляется до величины b_i . В результате дискретное представление аналогового сигнала преобразуется в цифровое, поскольку используется лишь конечное число возможных значений отсчетов.

Конечно, квантование сопряжено с известной ошибкой, в связи с чем и цифровое представление аналоговых сигналов оказывается лишь приближенным. Однако ошибка такого представления может быть сделана как угодно малой, если число уровней квантования достаточно велико. Если $x(t)$ случайный процесс, то его отсчеты — случайные величины. В этом случае ошибка квантования может быть оценена величиной своей дисперсии, которая во всех случаях, в первом приближении, обратно пропорциональна N^2 , так что можно ориентировочно принять

$$\sigma_k^2 = d \sigma_x^2 / N^2, \quad (1.63)$$

где σ_x^2 — дисперсия исходного процесса $x(t)$; d — коэффициент пропорциональности; d зависит от распределения вероятностей отсчета, размещения порогов и выбора уровней квантования. Для минимизации ошибки статистические характеристики случайных величин должны быть согласованы с параметрами квантователя.

Наиболее просто установить такое согласование при равномерном распределении вероятностей случайной величины. В этом случае интервалы квантования должны быть равными, а уровни квантования должны размещаться посередине между порогами. При таком оптимальном выборе параметров квантователя коэффициент d оказывается близким единице.

При нормальном распределении случайной величины равномерное расположение порогов перестает, естественно, быть оптимальным. Оказывается, что здесь пороги целесообразно распределять неравномерно, делая ширину интервалов квантования монотонно убывающей функцией плотности распределения вероятностей (меньшей близ нулевого уровня и большей по мере удаления от него). При этом для не слишком малого числа уровней квантования ($N > 10$) оказывается $d \rightarrow 2,72$.

Неравномерное квантование аппаратурно гораздо сложнее равномерного. Как правило, используют именно равномерный квантователь, подвергая предварительно исходный сигнал безынерционному нелинейному преобразованию (мгновенному компандированию), так чтобы имитировать оптимальное неравномерное квантование. При восстановлении аналогового сигнала по цифровому представлению (цифро-аналоговом преобразовании — ЦАП) используется экспандер — нелинейный усилитель с характеристикой вход-выход обратной реализуемой в компандере.

Итак, описанная процедура построения цифрового представления аналогового видеосигнала (аналого-цифрового преобразования — АЦП) сводится к последовательному выполнению дискретизации и квантования. Дисперсия общей ошибки АЦП для данного случая может быть найдена суммированием дисперсий (1.62) и (1.63):

$$\sigma_0^2 = \sigma_d^2 + \sigma_k^2. \quad (1.64)$$

Описанный алгоритм АЦП, именуемый ИКМ, хотя и является наиболее распространенным, но, тем не менее, не единственно возможный. Ниже будут отмечены и другие алгоритмы, в связи с чем возникает вопрос о сопоставлении показателей различных способов АЦП. Одним из основных таких показателей считается объем цифрового представления аналогового сигнала, определяемый выражением

$$G = F \log_2 N. \quad (1.65)$$

Объем измеряется в бит/с и определяет пропускную способность канала связи, который необходим для трансляции данного цифрового представления. Другие характеристики эффективности АЦП характеризуют точность, которая достигается при восстановлении аналогового сигнала по его цифровому представлению. В простых случаях указанная точность оценивается отношением дисперсии ошибки к дисперсии самого сигнала

$$\gamma^2 = \sigma_0^2 / \sigma_x^2. \quad (1.66)$$

Заслуживает внимания и другой, отличный от ИКМ, универсальный метод АЦП, именуемый Δ -преобразованием (модуляцией). В этом варианте на каждом такте работы АЦП предсказывается ожидаемое значение аналогового сигнала по координатам цифрового представления, полученным на предыдущих тактах. Из истинного значения поступающего на данном такте на вход АЦП аналогового сигнала, вычитается его предсказанное значение, эта разность квантуется, в результате чего и формируется очередная координата Δ -представления. Таким образом, алгоритм Δ -преобразования является рекуррентным. Известен целый ряд его разновидностей. Ниже ограничимся описанием лишь одного из них, который назовем *классическим*. В этом варианте используется простейшее предсказание вида

$$\hat{x}(t_k) = \sum_{l < k} \Delta b_l, \quad (1.67)$$

где в левой части предсказанное значение аналогового сигнала в k -й момент времени; Δ — некоторая константа; b_l — координаты Δ -представления. Суммирование ведется по всем $l < k$.

Далее вычисляется разность между истинным значением аналогового сигнала в точке t_k и предсказанным его значением в данной точке. Очередная координата Δ -представления

формируется путем квантования этой разности на двух уровнях; именно, принимается $b_k = +1$ или -1 в соответствии с тем положительна или отрицательна эта разность.

Восстановление аналогового сигнала по его классическому Δ -представлению осуществляется за счет фильтрации сглаживающим фильтром последовательности коротких импульсов, формируемых в тактовые моменты времени и имеющих площади, определяемые (1.67). В качестве сглаживающего используется фильтр нижних частот с близкой к П-образной АЧХ и частотой среза, совпадающей с верхней частотой спектра сигнала $x(t)$. Что касается частоты дискретизации $F_d = 1/q$, то при Δ -преобразовании ее приходится выбирать заметно большей, чем при ИКМ. Зато координаты могут принимать здесь всего два значения.

Объем классического Δ -представления

$$G = F_d [\text{бит/с}]. \quad (1.68)$$

Что касается точности такого представления, то она достаточно сложным образом зависит как от шага дискретизации q , так и от величины шага по уровню Δ . По поводу этих зависимостей ограничимся лишь пояснениями качественного характера.

На рис. 1.14 поясняется процедура формирования Δ -представления и выявляются некоторые особенности данного преобразования. Здесь представлен (сплошная линия) график отрезка аналогового сигнала $x(t)$, подлежащего цифровому преобразованию. На график нанесена сетка с шагом по оси времени (ось абсцисс) q , а по оси уровней (ось ординат) Δ . Примем для определенности, что в начальный момент времени истинное значение аналогового сигнала и его значение, предсказанное по предыдущим координатам Δ -представления, были весьма близки, но первое все-таки чуть больше второго, так что была выработана координата цифрового представления $+1$. (Координаты формируемого цифрового представления показаны в рамке под осью абсцисс в соответствии с моментами их появления.) Последовательность предсказанных значений \hat{x}_k , получаемых в соответствии с описанным выше алгоритмом Δ -преобразования, объединена ступенчатой линией. Восстановленный на приемной стороне случайный аналоговый сигнал представляет собой результат сглаживания фильтром ступенчатой кривой.

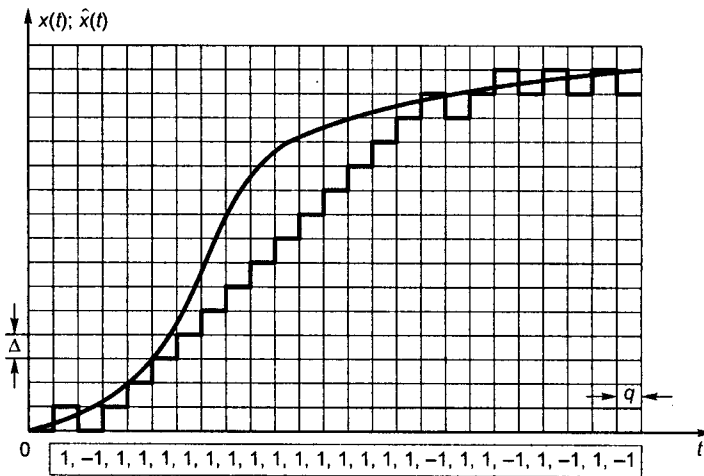


Рис. 1.14. К пояснению алгоритма Δ -преобразования

Из данных этого рисунка следует, что на пологих участках функции $x(t)$ аппроксимация колеблется вокруг ее истинных значений, удаляясь от них максимально на величину порядка Δ . На крутых участках изменения исходной функции аппроксимация отстает от нее, поскольку средняя крутизна изменения аппроксимирующей функции не может превосходить отношения Δ/q , в то время как $x(t)$, являясь случайным процессом, может, вообще говоря, с известной вероятностью достигать любой наперед заданной крутизны изменения. Из этих примеров следует, что при Δ -преобразовании всегда могут быть выделены интервалы времени двух типов: *интервалы слежения* (аппроксимация колеблется вокруг истинных значений) и *перегрузки* (аппроксимация отстает от изменений исходной функции). Соответственно рассматриваются и ошибки аппроксимации двух типов: шумы дробления, имеющие место в режиме слежения, и шумы перегрузки, возникающие в результате перегрузок.

Шумы дробления имеют порядок Δ , так что для их уменьшения следует уменьшать шаг Δ . Шумы перегрузки на интервалах перегрузки по порядку величины имеют тот же уровень, что и сама функция $x(t)$, но в среднем оказываются тем меньше, чем меньше вероятность наступления перегрузки. Описанный механизм возникновения перегрузок показывает, что вероятность их появления растет с уменьшением Δ , поскольку уменьшается допустимая крутизна изменения исходной функции. Таким образом, при изменении, например, росте Δ , проявляются два противоположных явления: уменьшаются шумы перегрузки и увеличиваются шумы дробления. Наоборот, если Δ уменьшать, то будут уменьшаться шумы дробления, но расти шумы перегрузки. Поскольку целью является достижение минимума суммарных шумов дробления и перегрузки, то существует некоторое оптимальное значение шага $\Delta = \Delta_{\text{опт}}$, при котором этот минимум достигается.

Оптимальное значение зависит как от статистических свойств преобразуемого аналогового сигнала, так и от его уровня. Поскольку эти параметры могут в процессе преобразования изменяться заранее неизвестным образом, то затруднительно было бы заранее выбрать и величину Δ . Тут, однако, на помощь приходит одно замечательное свойство Δ -представления, а именно тот факт, что вид последовательности его координат позволяет судить об оптимальности выбора шага Δ . Дело в том, что на интервалах слежения типично наличие большого числа знакоперемен в этой последовательности. Обычно за $+1$ в ней следует -1 и далее вновь $+1$ и лишь изредка встречаются два-три символа одного знака подряд. Наоборот, на интервалах перегрузки наблюдается последовательное повторение координат одного и того же значения ($+1$ при быстром нарастании преобразуемого аналогового сигнала и -1 при его быстром спаде). Соответственно шаг Δ должен выбираться так, чтобы второй случай встречался достаточно редко.

Указанное свойство используется для выбора оптимального шага следующим образом. Устанавливается, что значение Δ выбирается как определенная функция от числа знакоперемен в последовательности из определенного числа последних по времени координат цифрового представления. Эта функция и выбирается таким образом, чтобы получаемое значение шага было близким к оптимальному. Важно отметить, что поскольку указанная функция известна и на приемном конце радиолинии, то и там по получении Δ -представления можно определить текущее значение Δ ; никакого специального канала связи для передачи этих данных не требуется. Использование этого принципа позволяет автоматически изменять шаг при изменении тех или иных характеристик транслируемого аналогового сигнала, т.е. осуществлять компандирование.

Изложенное показывает, что по аппаратурной реализации Δ -преобразование гораздо проще ИКМ. Это определяется исключением многоуровневого квантования и исключительной простотой и надежностью процедуры компандирования. Наряду с этим достоинством

Δ -преобразованию присущи и определенные недостатки. Отметим, например, что это преобразование непригодно для нецентрированных случайных процессов, т.е. процессов с ненулевым математическим ожиданием. Вспомнив еще раз структуру алгоритма данного преобразования, можно заключить, что координаты представления отражают при этом преобразовании лишь изменения входного аналогового сигнала, но не его постоянную составляющую. Начальный уровень восстановленного сигнала на приемной стороне устанавливается здесь случайно и лишь изменения сигнала относительно этого уровня воспроизводятся правильно.

Любой сбой в работе радиoliniи приводит к новой случайной установке начального уровня, так что и достаточно низкочастотные составляющие процесса также не восстанавливаются достоверно. В результате Δ -преобразование оказывается пригодным лишь для процессов, энергетический спектр которых не содержит составляющих в некотором интервале частот с центром в нуле. Такие сигналы встречаются достаточно часто. Так, сигналы, передаваемые по ТЧ-каналу, содержат лишь составляющие в полосе частот от 0,3 до 3,4 кГц. Однако встречаются и видеосигналы, содержащие составляющие сколь угодно низких частот, как телевизионный сигнал. Их цифровое преобразование возможно методом ИКМ и невозможно способом Δ -модуляции.

Еще более существенным недостатком Δ -преобразования, является то обстоятельство, что по обеспечиваемой точности при одинаковых объемах представления, оно всегда уступает ИКМ и при том в тем большей степени, чем ближе энергетический спектр исходного аналогового процесса к П-образному. Для того чтобы обеспечить требуемую точность приходится, как правило, увеличивать объем Δ -преобразования по отношению к ИКМ. По этой причине на основных направлениях, определяющих требуемую пропускную способность каналов связи, используется, обычно ИКМ-преобразование (если необходимо универсальное АЦП). Δ -преобразование используется с целью упрощения аппаратуры лишь для отдельных сигналов, например, в каналах служебной связи, не влияющих существенно на общий баланс пропускной способности.

В заключение сделаем еще одно замечание по терминологии. Аббревиатура ИКМ происходит из термина импульсно-кодовая модуляция. Этот термин был введен достаточно давно, когда данный метод АЦП принято было сочетать с определенным способом модуляции. Сейчас его рассматривают совершенно независимо от используемого способа передачи получающегося цифрового представления, так что к модуляции он не имеет никакого отношения. Тем не менее аббревиатура ИКМ настолько укоренилась, что отказываться от нее сейчас нецелесообразно. Поэтому сочетание ИКМ рассматривают теперь просто как условное обозначение определенного способа АЦП, не прибегая к его расшифровке. Именно так мы и поступили выше, введя это обозначение без дополнительных пояснений. Те же соображения относятся и к термину Δ -модуляция.

На этом завершим краткий обзор вопросов, относящихся к АЦП видеосигналов. Более подробные данные по данным вопросам читатель найдет в [1.3], где имеются и дальнейшие литературные ссылки.

1.3.3. Цифровые представления радиосигналов

Известные обобщения теоремы Котельникова позволяют утверждать, что при определенных дополнительных условиях радиосигнал, занимающий суммарно полосу частот шириной F ($F/2$ в области положительных значений частоты и столько же в области ее отрицательных значений), может быть однозначно представлен своими отсчетами, взятыми с шагом по времени $q = 1/F$. Однако наличие указанных дополнительных условий, а также аппа-

ратурные осложнения, связанные с взятием отсчетов высокочастотных колебаний, приводят к тому, что практически АЦП подвергают не сами радиосигналы, а их те или иные видеоэквиваленты. Цифровые представления видеоэквивалентов являются одновременно и цифровыми представлениями исходных радиосигналов, поскольку позволяют их восстановить с любой требуемой точностью. Если же, как это часто бывает, нас интересует не сам радиосигнал, а лишь некоторые его параметры, то восстановление сигнала вообще не производят, сразу вычисляя эти параметры по цифровым представлениям его видеоэквивалентов.

В системах цифровой связи чаще всего приходится находить законы изменения огибающей и фазы радиосигнала. Покажем, как эти параметры могут быть определены по цифровым представлениям ПВЭ и КК.

Поскольку ПВЭ $v(t)$ (см. (1.57)) имеет ограниченный спектр, он может быть представлен рядом Котельникова:

$$v(t) = \sum_{k=-\infty}^{\infty} v_k \left(\frac{k}{2\Delta f} \right) \frac{\sin[2\pi\Delta f(t - k/(2\Delta f))]}{2\pi\Delta f(t - k/(2\Delta f))} = \sum_{k=-\infty}^{\infty} v_k \frac{\sin \pi(\tau - k)}{\pi(\tau - k)}, \quad (1.69)$$

где $v_k = v\left(\frac{k}{2\Delta f}\right)$ и $\tau = 2\Delta f t$.

Покажем, как по ПВЭ можно определить огибающую и фазу радиосигнала. Из теории сигналов известно, что понятие огибающей и полной фазы сигнала может быть распространено и на видеосигналы, путем использования преобразования Гильберта. Для функции $x(t)$ преобразованием Гильберта называют функцию

$$\hat{x}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(t_1)}{t - t_1} dt_1. \quad (1.70)$$

Подставляя (1.69) в (1.70), найдем преобразование Гильберта от ПВЭ:

$$\hat{v}(t) = \sum_{k=-\infty}^{\infty} v_k \frac{\sin^2[0,5\pi(\tau - k)]}{0,5\pi(\tau - k)}. \quad (1.71)$$

Отметим, что в l -й тактовой точке, т.е. при $t = l/(2\Delta f)$ все члены в выражении (1.69) для ПВЭ, кроме l -го, обращаются в нуль. Что же касается преобразования Гильберта от ПВЭ, то на него это свойство не распространяется: все члены с номерами k такими, что $l - k$ нечетно дают, вообще говоря, конечный вклад в (1.71).

Как это следует из теории сигналов, огибающая и полная фаза ПВЭ могут быть выражены следующими соотношениями:

$$A(t) = \sqrt{v^2(t) + \hat{v}^2(t)}; \quad (1.72)$$

$$\theta(t) = \text{arccoor}[\hat{v}(t), v(t)]. \quad (1.73)$$

В последнем равенстве введена функция $\text{arccoor}(x, y)$, определяемая как угол на плоскости между горизонтальной осью и вектором с координатами x и y вдоль вертикальной и горизонтальной осей соответственно. Можно было бы записать этот угол традиционно, через arctg , но при этом потребовалась бы достаточно громоздкая запись, учитывающая знаки координат, поскольку сам arctg определяет лишь значение угла по модулю π . В нашем случае, поскольку средняя частота для ПВЭ составляет $\Delta f/2$, имеем $\theta(t) = \pi\Delta t + \varphi(t)$, так что, используя (1.72), получим окончательно:

$$\varphi(t) = \text{arccoor}[\hat{v}(t), v(t)]. \quad (1.74)$$

Квадратурные координаты действительно позволяют восстановить радиосигнал. В частности, его огибающая:

$$A(t) = \sqrt{Q^2(t) + I^2(t)} \quad (1.75)$$

и фаза

$$\varphi(t) = \operatorname{arctg}[I(t); Q(t)]. \quad (1.76)$$

Что касается преобразования радиосигналов в цифровую форму, то для этой цели используют следующую процедуру: вначале одним из описанных выше способов образуют аналоговые видеосигналы, эквивалентные данному радиосигналу, а затем их оцифровывают. Получающиеся цифровые последовательности и рассматриваются как цифровое представление радиосигнала. Эти последовательности могут быть пересчитаны по приведенным выше формулам в последовательности отсчетов огибающей и фазы радиосигнала. При этом, если используются КК, то для пересчета должны использоваться выражения (1.75), (1.76), а для ПВЭ — выражения (1.72) и (1.76).

Отметим, что совершенно не обязательно вычислять отсчеты огибающей и фазы в те же моменты времени, для которых выявлены отсчеты видеоеквивалентов сигнала. Действительно, видеоеквиваленты определяются своими отсчетами однозначно в любой момент времени рядами Котельникова вида (1.69). Таким образом, данные отсчеты могут быть пересчитаны в любые другие, сдвинутые по времени, или, даже, взятые с другой частотой. Процедура изменения частоты отсчетов и сдвига их временных положений широко используется при цифровой обработке сигналов. При этом преобразования, связанные с повышением частоты отсчетов, именуется *интерполяцией*, а с ее уменьшением — *децимацией*.

1.3.4. Базовые методы цифровой модуляции

Под *модуляцией* понимают процесс преобразования подлежащего передаче сообщения в сигнал, который может быть передан по каналу связи. В соответствии с задачами этой книги, будем здесь рассматривать только те случаи, когда сообщение представлено в виде цифрового видеосигнала. Именно в связи с этим рассматриваемые ниже методы модуляции называются *цифровыми*. Часто цифровую модуляцию именуют специальными терминами: телеграфия или манипуляция; последний термин чаще всего используют в аббревиатурах, поскольку аббревиатуры терминов модуляция и манипуляция совпадают. Ниже рассмотрим только тот типовой для радиосвязи случай, когда манипулируемый сигнал является аналоговым радиосигналом. В качестве переносчика при формировании радиосигнала используется гармоническое колебание несущей частоты $u(t) = A \cos(\omega_n t + \varphi)$. Процесс манипуляции заключается в изменении одного или нескольких параметров этого колебания (амплитуды, частоты или фазы) в соответствии с очередным поступающим символом сообщения. В простейшем случае, если такт поступления символов сообщения составляет T_c , то при манипуляции с тем же тактом происходят и изменения параметров манипулируемого колебания. При этом число градаций манипулируемого параметра равно значности символов поступающего сообщения, а номер выбираемой градации определяется значением очередного символа. Количество используемых градаций параметров несущей существенно влияет на основные характеристики осуществляемой передачи информации. Поэтому в общем случае перед подачей на модулятор сообщение подвергают специальному обратимому безызбыточному преобразованию, изменяющему его значность, а следовательно, и длительность такта (с T_c , например, T_3). На приемной стороне радиолинии принятый на каждом тактовом интервале длительностью T_3 сигнал анализируется (демодулируется) с целью опреде-

ления номера выбранной на передаче градации параметра, а следовательно, и значения транслируемого символа преобразованного сообщения. Далее осуществляется обратное преобразование сообщения, восстанавливающее его исходную форму, в которой оно и транслируется пользователю.

Иногда удобно рассматривать процесс модуляции несколько иначе. Именно, с каждым возможным значением символа преобразованного сообщения связывают тот или иной радиосигнал или в общем случае группу сигналов, каждый длительностью T_s . При поступлении на вход модулятора очередного символа преобразованного сообщения модулятор выбирает сигнал из той группы, которая соответствует значению этого символа и передает его в канал связи. Выбранные таким образом радиосигналы длительностью T_s именуется *элементарными сигналами* (ЭС), или элементарными посылками. Процесс демодуляции, осуществляемый на приемной стороне радиолинии, рассматривается как определение номера ЭС или группы, к которой принадлежит ЭС, поступивший по радиолинии на данном такте. Тем самым определяется и значение символа сообщения, транслировавшегося по радиолинии на этом такте. Такое представление процесса модуляции и демодуляции фактически не отличается от предыдущего: ведь каждый радиосигнал может быть представлен как гармоническое колебание с теми или иными параметрами, так что последовательность ЭС, выбираемых согласно с поступающим сообщением, представляет собой гармоническое колебание с изменяющимися в соответствии с этим сообщением параметрами.

Следует подчеркнуть, что при изменениях частоты или фазы однозначное разграничение этих параметров становится невозможным. Фактически по радиосигналу однозначно определяется только полная фаза $Q(t) = \int_0^t \omega_n(t) dt + \varphi(t)$, а выбор значений частоты $\omega_n(t)$ и фазы $\varphi(t)$ перестает быть однозначным: можно практически произвольно выбрать один из параметров, после чего в соответствии со значениями $Q(t)$ определится другой.

Мгновенная частота манипулированного радиосигнала $\omega(t) = \frac{d}{dt} Q(t)$ является случайным процессом, поскольку зависит от последовательности символов сообщения, которые всегда в технике связи носят случайный характер (в противном случае не было бы необходимости в их передаче). Обозначим среднее значение этой частоты через ω_0 . Удобно записывать манипулированный радиосигнал именно с использованием данной частоты в виде $u_c(t) = A(t) \cos[\omega_0 t + \psi(t)]$. В соответствии с определением средней частоты ω_0 , $\overline{\frac{d}{dt} \psi(t)} = 0$, так что фаза $\psi(t)$, приведенная к средней частоте, не может преимущественно нарастать или спадать: спад и нарастание встречаются равновероятно, так что фаза колеблется вокруг некоторого постоянного значения. Ни при каком другом выборе частоты колебания этот факт не имеет места: фаза относительно любой другой частоты будет в среднем постоянно нарастать или уменьшаться.

К базовым методам манипуляции будем относить те, при которых за время элементарной посылки параметры радиосигнала остаются постоянными, изменяясь скачком в тактовых точках. В зависимости от того, какой именно параметр манипулируемый, различают амплитудную (АМ), частотную (ЧМ) или фазовую (ФМ) манипуляции. Условимся сразу вслед за аббревиатурой метода манипуляции цифрой указывать число используемых градаций параметра. Так, запись ФМ4 означает, что речь идет о фазовой манипуляции с четырьмя градациями фазы (обычно $0, \pi/2, \pi$ и $3\pi/2$).

В качестве примера на рис. 1.15 условно показаны осциллограммы радиосигналов при передаче цифровой информации методами АМ2 (с пассивной паузой), ФМ2 и ЧМ2 при трансляции сообщения, состоящего из двух двоичных цифр (транслируемые цифры сообщения приведены над посылками). Условность изображения связана с представлением несущей частоты (заполнения посылок). Как правило, эти частоты столь высоки, а их периоды соответственно, столь малы по сравнению с T_s , что на реальных осциллограммах радиосигналов заполнение посылок выглядит сплошным. Поскольку нам необходимо проследить, как меняется заполнение от посылки к посылке, на приведенных рисунках заполнение посылок представлено без соблюдения масштаба.

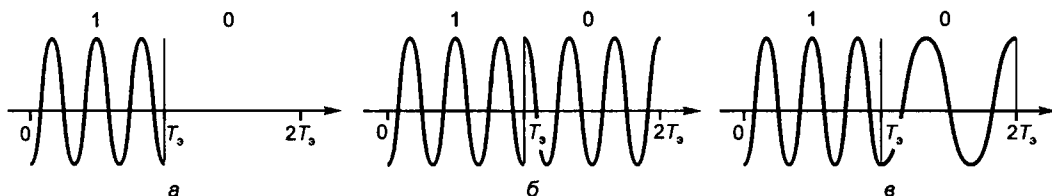


Рис. 1.15. Сигнал АМ2 с пассивной паузой (а), ФМ2 (б) и ЧМ2 (в)

Как увидим ниже, фазовая манипуляция обладает рядом достоинств. Вследствие чего в настоящее время основанные на ней методы передачи информации чаще всего и используются. Однако ей присущ один специфический недостаток, именуемый «обратной работой». Суть соответствующей проблемы удобно пояснить на примере ФМ2. При формировании соответствующего сигнала на передающем конце радиолинии некоторая фаза несущей частоты принимается за нулевую, и ей ставится в соответствие, например, символ сообщения «0». Тогда несущей в противоположной фазе ставится в соответствие символ «1». На приемном конце радиолинии, если принимаемый сигнал в достаточной степени превосходит шумы, можно зафиксировать появление ЭС с двумя противоположными фазами. Здесь нет никаких сведений относительно того, какую из этих фаз следует считать нулевой. Поэтому демодулятор может равновероятно принять за нулевую любую из двух фаз принимаемых ЭС; если один из таких выборов приведет к правильной демодуляции транслируемого сообщения, то второй приведет к тому, что будет принято обратное сообщение: всюду вместо «1» будут фигурировать «0», и наоборот. Это явление и называют *обратной работой*. Указанное название сохраняется и за аналогичным явлением, возникающим и при m -ичной фазовой манипуляции. И здесь любая из фаз, поступающих на демодулятор ЭС, может быть принята им за нулевую, что, как правило, приводит к специфическому искажению всего демодулированного сообщения. Характер соответствующих искажений несколько более сложен, чем простая замена символов. В частности, из-за ошибочного выбора на приеме начальной фазы каждое переданное сообщение может трансформироваться в одно из m возможных принятых; только одно из них признается правильным, остальные будем по-прежнему именовать обратными.

Отметим, что для устранения обратной работы требуются дополнительные сведения весьма небольшого объема. Действительно, за время непрерывной работы радиолинии достаточно один раз передать дополнительное сообщение, позволяющее выбрать правильную опорную фазу из m возможных. Объем такого сообщения не превосходит нескольких бит. Однако его передача всякий раз, когда по тем или иным причинам правильное значение опорной фазы в демодуляторе оказывается потерянным, вызывает определенное усложнение алгоритмов организации и поддержания связи.

Часто поступают следующим образом: на входе модулятора ФМ устанавливается дополнительный преобразователь сообщений (кодер), а на выходе демодулятора соответственно декодер. Кодирование осуществляется таким образом, чтобы прямая и обратные комбинации на выходе кодера соответствовали одному и тому же сообщению на его входе. Тогда, если в канале связи возникнет обратная работа, на содержании декодированного сообщения это не отразится. Ясно, что такое кодирование требует увеличения числа символов в кодовых комбинациях (на выходе кодера) по сравнению с их числом в сообщении. Поскольку число кодовых комбинаций должно превосходить число сообщений ровно в m раз, то это означает, что число m -ичных символов кодовой комбинации должно превосходить число таких же символов сообщения в точности на один символ. Важно отметить, что этот факт не зависит от числа символов в сообщении; во всех случаях достаточно при образовании кодовой комбинации добавить к сообщению один символ.

Как же конкретно может выполняться такое кодирование? Очевидно, что оно должно использовать те параметры кодовой комбинации, которые не изменяются при возникновении обратной работы. При ФМ набор таких параметров исчерпывается величинами изменений фазы в каждой тактовой точке, т.е. значениями разности фаз каждой посылки ЭС относительно предыдущей. Именно этим разностям должны ставиться в соответствие символы передаваемого сообщения. Так, при двоичной передаче можно использовать следующий способ кодирования: если на очередном тактовом интервале должен быть передан символ «0», то на нем передается ЭС с тем же значением фазы, которое имела предыдущая посылка. Если же необходимо передать «1», то фаза ЭС относительно предыдущей посылки изменяется на π . Такое кодирование называют относительным. Для ФМ4 относительное кодирование может выполняться, например, в соответствии с данными табл. 1.2. Аналогично осуществляется относительное кодирование и при ФМ большей значности.

Таблица 1.2. Параметры относительного кодирования для ФМ4

Передаваемый блок символов сообщения	00	01	10	11
Сдвиг фазы в тактовой точке	0	$\pi/2$	$-\pi/2$	π

С учетом изложенного, передача с относительным кодированием возможна, если уже был передан предыдущий ЭС, относительно которого и, в соответствии с очередными символами сообщения, выбирается следующий ЭС. Поэтому реально в приемнике начинается правильное декодирование очередных символов сообщения только после приема предыдущей ЭС. Это соответствует требованию увеличения числа символов в кодовой комбинации на единицу по сравнению с числом символов сообщения. Если при использовании относительного кодирования в радиолинии произойдет нарушение приема, то декодирование начнется только со второго из принятых после такого нарушения ЭС: первый не даст правильного сообщения. Очевидно, что возникающее в связи с этим удлинение нарушения связи на один ЭС не является существенным, зато не требуется каких-либо специальных мер для восстановления связи после ее нарушения.

Фазовая манипуляция в сочетании с относительным кодированием получила специальное название: относительная фазовая манипуляция (ОФМ). Тот факт, что такое сочетание рассматривается как новый метод манипуляции не случайно. Дело в том, что добавление к модулятору ФМ относительного кодера придает ему принципиально новые свойства. В частности, сигналы ОФМ допускают демодуляцию в тех условиях, когда демодуляция сигналов ФМ принципиально невозможна (более подробно см. гл. 3).

Глава 2

ТЕОРИЯ ИНФОРМАЦИИ И ДОСТИЖИМАЯ ЭФФЕКТИВНОСТЬ СИСТЕМ ЦИФРОВОЙ РАДИОСВЯЗИ

2.1. Мера количества информации и производительность источника

Теория информации посвящена рассмотрению проблем передачи информации в весьма общем контексте. Здесь заданными считаются только параметры канала связи, такие как частотная характеристика, мощности сигнала (пиковая и средняя), характеристики помех и т.п. Задача рассмотрения состоит в оценке максимально возможной скорости передачи информации по такому каналу при выборе наилучших из возможных сигналов, методов их передачи и приема. Именно передача информации представляет собой основную проблему техники связи. Однако до появления теории информации само понятие информации вводилось лишь на интуитивном, качественном уровне. Важная заслуга теории — уже само установление количественной меры информации, позволившее формулировать задачи техники связи количественно. Очевидно, что невозможно ввести такую количественную меру информации, которая полностью учитывала все ее аспекты и удовлетворила бы любых пользователей, независимо от задач, для решения которых они прибегают к этой мере.

Трудно ожидать, например, что будет найдена мера, учитывающая и эстетические впечатления слушателя классической музыки, и объем сведений, содержащихся в отчете о деятельности некоего банка. Теория информации и не ставит перед собой задачи поиска такой универсальной меры. Поскольку эта теория нацелена на анализ систем связи, достаточной для нее является мера, учитывающая только те характеристики информации, которые существенны для таких систем. В качественном плане можно утверждать, что система связи призвана транслировать получателю, подключенному к выходу линии связи, состояния источника сообщений, подключенного к ее входу. Для линии связи существенны число возможных состояний источника, частота смены этих состояний и их статистика. Именно перечисленные характеристики и учитываются мерой количества информации, вводимой в теории информации.

Такая мера позволяет рассмотреть процессы преобразования количества информации в тех или иных устройствах, входящих в состав систем связи, в результате чего устанавливаются основополагающие законы, определяющие связь между характеристиками канала связи и количеством информации, которая может быть по нему передана. Эти законы, похожие, в известном смысле, на законы сохранения энергии, служат фундаментом теории связи, придающим ей черты не только технической, но и научной дисциплины. То же, в определенной мере, относится и к рассмотренной в следующей главе теории оптимального приема сигналов.

2.1.1. Мера информации

В теории информации исследуются два существенных объекта: источник и получатель информации. Важнейшей особенностью информационного сигнала на выходе источника считается его в точности непредсказуемый, случайный для системы связи характер. В противном случае использование линии связи для трансляции такой информации теряло бы смысл. Работа источника рассматривается как выбор в соответствии с заданными статистическими закономерностями, подлежащего передаче сообщения из известного ансамбля возможных сообщений. Предполагается, что на приемной стороне линии связи известны заранее как сам ансамбль, так и используемая при выборе статистика. Эти сведения являются для приемника априорными. Они не позволяют однозначно определить сообщение, которое будет выбрано источником, в связи с чем до получения данного сообщения получатель находится в состоянии известной неопределенности. Эта неопределенность будет снята получением сообщения от источника.

В теории информации вводится рациональная количественная мера неопределенности, соответствующей тем или иным априорным данным. Эта мера именуется *энтропией* и обозначается символом H . Далее принимается, что количество информации, доставляемой получателю наблюдением выбранного источником сообщения, и определяется именно величиной H . Иначе говоря, до получения сообщения у получателя существовала неопределенность относительно исхода опыта по выбору сообщения источником. Количественно она измерялась величиной H . После проведения опыта (получения сообщения) указанная неопределенность была снята. Количество информации в сообщении и характеризуется объемом снятой им неопределенности H .

Остановимся на вопросе об установлении количественной меры неопределенности H . Итак, пусть имеется некоторая схема статистических испытаний, полное множество несовместных исходов которой содержит N событий. Пусть известны вероятности этих исходов Q_j ; $j = 1, 2, \dots, N$. Как измерить энтропию данной схемы? Для того чтобы обосновать соответствующую формулу, автор теории информации, американский ученый Клод Шеннон, поступил следующим образом: вначале он сформулировал несколько требований, которые разумно предъявить к количественной мере неопределенности, а затем показал, что фактически существует только одно математическое выражение, удовлетворяющее этим требованиям.

Перечислим эти требования.

Требование 1. При фиксированном числе исходов N неопределенность H должна достигать максимума, когда вероятности всех исходов одинаковы (т.е. $Q_j = 1/N$). Действительно, очевидно, что именно в этом случае вероятность правильного предсказания исхода опыта минимальна.

Требование 2. При равновероятных исходах неопределенность должна возрастать с ростом N . Действительно, с ростом N вероятность правильного предсказания исхода опыта падает.

Требование 3. Неопределенность должна быть равна нулю в том и только в том случае, когда исход опыта предрешен (один из исходов является достоверным событием). Действительно, в этом случае исход опыта заранее известен.

Требование 4. Это требование более сложно, чем предыдущие и относится к варианту, когда реализуется сложная схема испытаний, объединяющая две более простые; естественно, что в этом случае неопределенность сложной схемы должна быть соответствующим образом связана с неопределенностями образующих ее схем.

Именно, оказывается достаточным рассмотреть двухшаговую схему испытаний, когда статистика испытаний на втором шаге, зависит от исхода испытаний первого шага. Общая

энтропия схемы должна быть суммой энтропии, соответствующей первому шагу и взвешенной суммы энтропий второго шага.

Можно показать, что этим требованиям удовлетворяет единственное выражение

$$H = -c \sum_{j=1}^N Q_j \lg Q_j. \tag{2.1}$$

Здесь $c > 0$ — произвольная константа, определяющая на самом деле используемые единицы измерения неопределенности. Эта величина всегда может быть заменена соответствующим выбором основания логарифма, входящего в эту формулу. Везде ниже полагается $c = 1$, а основание логарифма равно 2, так что

$$H = -\sum_{j=1}^N Q_j \log_2 Q_j. \tag{2.1a}$$

В настоящей главе для сокращения записи прямое указание на основание логарифма будет опускаться.

Последняя формула показывает, что $H = 1$ соответствует схеме с двумя равновероятными исходами. Таким образом, эта схема реализует единицу измерения неопределенности. Единица именуется двоичной, или «битом».

Легко проверить, что полученное выражение действительно удовлетворяет приведенным требованиям. Начнем с простейшего примера схемы испытаний с двумя возможными исходами. Пусть вероятность первого исхода Q_1 , а второго $Q_2 = 1 - Q_1$. Зависимость энтропии такой схемы от Q_1 представлена на рис. 2.1.

Из рисунка следует, что энтропия максимальна при равенстве вероятностей возможных исходов. Можно доказать, что это справедливо и для схем с любым другим числом исходов.

Далее легко установить, что для равновероятных исходов

$$H = \log N, \tag{2.2}$$

так, что H действительно растет с ростом N . Если один из возможных исходов является достоверным событием, то вычисление по формуле (2.1a) дает $H = 0$.

Остановимся теперь более подробно на случае, когда одновременно реализуются две схемы испытаний в общем случае статистически взаимосвязанные. Для того, чтобы различать энтропии, относящиеся к различным испытаниям, будем снабжать их индексами, указывающими условные наименования этих схем. Обозначим первую схему через x , вторую через y , а сложную схему, заключающуюся в одновременной реализации обеих испытаний, через xy . Индексом n будем обозначать номер исхода первой схемы, индексом k — второй. Соответственно Q_n и Q_k — вероятности исходов первой и второй схем, а $Q_{n,k}$ — вероятности исходов сложной схемы. Тогда, используя выражение (2.1a), получаем

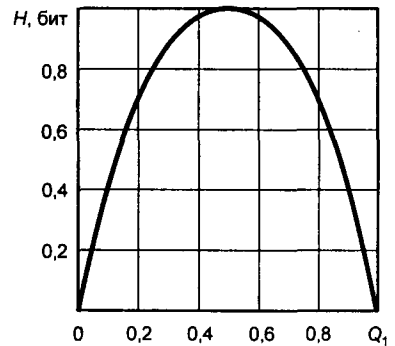


Рис. 2.1. Энтропия при двоичных испытаниях

$$H_{xy} = -\sum_{n,k} Q_{n,k} \log Q_{n,k} = -\sum_{n,k} Q_{n,k} (\log Q_n + \log Q_{k/n}) = H_x + \left(-\sum_n Q_n \sum_k Q_{k,n} \log Q_{k/n} \right),$$

где через $Q_{k/n}$ обозначена условная вероятность появления k -го исхода события y (y_k) при условии, что произошло x_n . Последняя скобка в этом выражении по определению именуется

условной энтропией $H_{x/y}$ схемы x по схеме y ; она показывает, какова в среднем неопределенность схемы x , если известен исход испытаний для схемы y .

Итак,

$$H_{y/x} = -\sum_n Q_n \sum_k Q_{k/n} \log Q_{k/n} \quad (2.3)$$

и

$$H_{x,y} = H_x + H_{y/x}. \quad (2.4)$$

Последнее выражение свидетельствует, в частности, о выполнении требования (2.4). По смыслу ясно, что должно выполняться соотношение

$$H_{y/x} \leq H_y, \quad (2.5)$$

поскольку неопределенность исхода некоторого испытания не может увеличиться от того, что станет известным результат любого другого опыта. Это утверждение действительно справедливо и может быть математически доказано. При этом равенство в приведенном выражении оказывается возможным лишь в случае, когда испытания x и y статистически независимы.

Сопоставляя последнее соотношение с (2.4), находим, что

$$H_{x,y} \leq H_x + H_y, \quad (2.6)$$

причем равенство в последнем соотношении имеет место, только если x и y статистически независимы.

2.1.2. Производительность источника информации

Остановимся теперь на описании источников информации. Источник цифровой информации отличается от описанной выше схемы статистических испытаний тем, что он проводит такого рода испытания последовательно во времени, одно за другим и выдает в канал последовательность номеров выпавших исходов. Такие последовательности будем именовать *цифровыми цепочками*, или просто цепочками. В общем случае между последовательно проводимыми источником испытаниями существует статистическая связь, следствием которой является статистическая зависимость между отдельными символами цепочки. Как в этих условиях воспользоваться приведенным определением энтропии для оценки информационной производительности источника?

Достаточно общий подход заключается в следующем: рассмотрим набор всевозможных цепочек источника достаточно большой длины T как набор возможных несовместных исходов некоторой укрупненной схемы статистических испытаний. Зная статистические свойства источника, принципиально можно подсчитать вероятности появления каждого такого исхода (цепочки) и далее в соответствии с (2.1a) — энтропию H_m этой схемы. Как правило, с ростом временного интервала между цифрами цепочки статистическая связь между ними убывает, так что можно ожидать, что отношение H_m/T с ростом T стремится к некоторому пределу. Этот предел и называют *производительностью источника*:

$$H_f = \lim_{T \rightarrow \infty} H_m / T. \quad (2.7)$$

Размерность определенной таким образом производительности источника бит/с.

Если обозначить число цифр в цепочке через M , а энтропию соответствующей укрупненной схемы испытаний через H_M , то количество информации источника, приходящейся на цифру (бит/цифру), можно определить выражением

$$H_f = \lim_{M \rightarrow \infty} H_M / M. \quad (2.8)$$

Очевидно, что в простейшем случае, когда каждая очередная цифра цепочки получается в результате независимых одинаковых (в смысле статистики) испытаний, количество информации, даваемой источником на цифру, $H_f = H_1$.

Достаточно простое выражение для количества информации, даваемой источником, может быть получено для случая, когда его работа описывается цепью Маркова. В этом случае статистическая связь между последовательно выдаваемыми источником цифрами описывается следующим образом. Вводится понятие состояния источника, которое зависит от нескольких, например m , последних выданных источником цифр. Если число различных значений, из которых каждый раз выбирается очередная цифра, равно n , то общее число состояний источника $L = n^m$.

Будем полагать, что вероятность появления той или иной цифры на выходе источника на данном такте его работы зависит от состояния источника на момент наступления этого такта. Обозначим эти вероятности через $W_{j,i}$, где j — номер состояния источника (от 1 до L), а i — номер выдаваемой им цифры (от 1 до n). Что касается самих состояний, то вероятность их наступления в стационарном режиме работы источника обозначим через V_j . Тогда количество информации, даваемой источником, определится в соответствии с приведенными выше основными свойствами энтропии соотношением

$$H_f = - \sum_j V_j \sum_i W_{j,i} \log W_{j,i} . \quad (2.9)$$

Работу источника можно рассматривать не только как выдачу одну за другой цифр сообщения, но и как процесс смены состояний, причем по последовательности цифр однозначно определяется последовательность состояний, и наоборот. Поэтому выражение (2.9) останется справедливым, если под $W_{j,i}$ понимать вероятность перехода за один такт из состояния j в состояние i . Если источник работает в постоянном такте, так что он выдает одну цифру каждые T_1 секунд, то его производительность составляет

$$H_t = H_f / T_1 . \quad (2.10)$$

2.2. Информация при детерминированных преобразованиях сообщений и основная теорема для канала без шумов

2.2.1. Детерминированные преобразования сообщений

Над выходными цепочками источника могут осуществляться функциональные (неслучайные) преобразования. Устройства, осуществляющие такие преобразования, будем именовать преобразователями. На рис. 2.2 показано соединение источника с преобразователем. Последовательность выходных цифр источника обозначена через a_k , а преобразователя — через b_l , где k и l — номера цифр в соответствующих последовательностях. Различия в обозначении номеров цифр обеих последовательностей подчеркивает то обстоятельство, что такт работы преобразователя может отличаться от такта работы источника.

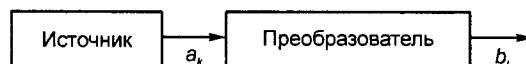


Рис. 2.2. Соединение источника с преобразователем

Аналогично, отличие обозначения самих цифр обеих последовательностей указывает на то, что число их возможных значений для этих последовательностей в общем случае также различно.

Выход преобразователя с подключенным к нему источником может рассматриваться как новый, составной источник, определенным образом связанный с основным. Если преобразователь различным входным цепочкам всегда ставит в соответствие различные выходные цепочки, т.е. если осуществляемое им преобразование является обратимым, то число различных входных и выходных цепочек достаточно большой длины T , также как и наборы вероятностей их появления одинаковы. Поэтому H_i для основного и составного источников одинаковы. Если же преобразователь осуществляет необратимое преобразование, т.е. преобразует несколько входных цепочек в одну выходную, то, в соответствии со свойствами энтропии, H_i для основного источника меньше, чем для составного. Итак, если производительности основного и составного источников обозначить через H_i^a и H_i^b соответственно, то можно утверждать, что

$$H_i^a \geq H_i^b, \quad (2.11)$$

причем равенство имеет место, только если преобразование обратимо. Следует отметить, что это выражение справедливо только для производительности источника (в среднем в единицу времени); аналогичное соотношение для количества информации на цифру было бы, вообще говоря, неверным, поскольку такты выдачи цифр на выходах основного и составного источников могут быть различными. В частности, можно пытаться так подбирать вид обратимого преобразования, чтобы величина H_i повышалась. Процедура, часто используемая для достижения такого эффекта, носит название *укрупнения алфавита* и сводится к тому, что несколько следующих подряд цифр последовательностей объединяются в одну укрупненную. Для того, чтобы такое преобразование было обратимым, число возможных значений укрупненных цифр должно быть соответствующим образом увеличено по сравнению с числом значений исходных. Так, если число последних составляет n и при укрупнении формируются группы по k таких цифр, то число значений укрупненных цифр должно составлять $n_v = n^k$.

Введем теперь еще одно понятие — *объем сообщений источника*. Эта величина может вычисляться по отношению к цифре сообщения (B_f) или за единицу времени (B_t) и определяется соотношениями:

$$B_f = \log n; \quad B_t = \frac{1}{T} \log n. \quad (2.12)$$

Как следует из последних соотношений, объем сообщения вычисляется независимо от статистических свойств последовательностей, переносящих эти сообщения. В соответствии со свойствами энтропии, он всегда больше или равен соответствующим характеристикам источника — объема информации на символ или производительности источника. Понятие объема сообщения вводится в связи с тем, что при передаче по каналу связи, существующие внутри сообщения статистические связи, как правило, не используются, так что требования к каналу определяются именно объемом транслируемого сообщения. В качестве единицы измерения объема сообщения сохраняют бит, хотя смысл этой величины здесь несколько меняется: она оценивается теперь просто количеством переданных чисел (в пересчете на двоичные числа), независимо от статистики их поступления. Объем сообщения совпадает с количеством информации в нем, вычисленным в предположении, что появление всех цифр равновероятно, и статистическая связь между цифрами, передаваемыми на различных тактовых интервалах, отсутствует.

Под *избыточностью сообщения* q понимают величину

$$q = \frac{B_t - H_t}{B_t}. \quad (2.13)$$

Определенная таким образом величина показывает, какая часть объема сообщения является как бы бессодержательной, излишней, возникшей лишь в связи с неудачным представлением информации в выходной последовательности источника. Преобразования, направленные на сокращение избыточности, носят название *кодирования источника*. Такое кодирование облегчает требования к каналу связи и поэтому признается актуальной технической задачей. Одним из путей ее решения считается упомянутая выше процедура укрупнения алфавита. Идея здесь заключается в том, что по мере укрупнения цифр статистические связи между ними, как правило, ослабляются.

Рассмотрим, например, возможности сокращения избыточности при передаче русского текста. Начнем с побуквенной передачи. При этом для передачи одного знака (буквы, пробела, знака препинания) требуется около 5,4 бит/знак. Можно укрупнить алфавит, перейдя от передачи букв к передаче слов. Для этого можно ограничиться набором из, например, 10–15 тысяч слов. С учетом их грамматических форм примем, что количество возможных значений укрупненных чисел составляет 64 000, так что для передачи каждого такого числа потребуется около 16 бит. Статистический анализ русских слов показывает, что в среднем по письменному тексту они содержат 6,3 букв на слово, так что при побуквенной передаче для передачи слова потребовалось бы порядка 39,4 бит/слово. Таким образом, при переходе от побуквенной передачи к пословной объем сообщения может быть сокращен в 2,5 раза. Естественно, что существенно сокращается избыточность сообщения. Однако она устраняется еще далеко не полностью. И дело здесь не только в том, что и между словами в их последовательности сохраняется определенная статистическая связь.

Важно, что процедура укрупнения алфавита разрушает только статистические связи между последовательно передаваемыми цифрами сообщения, но не учитывает различий вероятностей появления (частот встречаемости) различных укрупненных цифр. Так, частоты появления различных русских слов сильно различаются: такие слова как союзы и предлоги, также как и общеупотребительные слова встречаются гораздо чаще остальных. Основным способом сокращения избыточности, вызванной этим обстоятельством, считается вариация длительности тактов передачи укрупненных чисел: для передачи тех их значений, которые встречаются чаще других, отводятся более короткие такты. При этом кодирование источника должно осуществляться таким образом, чтобы сведения о длительности очередного такта содержались в самой кодированной последовательности (не должны были бы передаваться отдельно).

Чаще всего значения цифр на выходе источника записываются комбинациями двоичных чисел, так что его выходной сигнал представляет собой двоичную последовательность. Коды отдельных цифр должны быть выбраны так, чтобы границы кодовых комбинаций сразу же обнаруживались. Этим свойством обладают, например, префиксные коды. Если комбинации этих кодов выстроить в ряд, так чтобы более короткие предшествовали более длинным, то для любой конкретной комбинации ни одна из предшествующих ей не будет совпадать с ее начальной частью (префиксом). Ясно, что в этом случае разделение кодовых комбинаций осуществляется однозначно и без дополнительных указаний.

2.2.2. Основная теорема для канала без шумов

Указанные способы сокращения избыточности позволяют приблизить объем сообщения к количеству информации в нем. Более того, оказывается, что за счет соответствующих обра-

тимых преобразований избыточность может быть сделана сколь угодно близкой к нулю. Это утверждение составляет содержание основной теоремы теории информации для канала без шумов. К ее изложению мы и перейдем.

Канал связи без шумов описывается данными о тех последовательностях цифр, которые могут быть по нему переданы, т.е. которые можно различить на его выходе. Эти данные могут быть представлены в виде набора ограничений на последовательно передаваемые цифры, накладываемых каналом или прямым указанием допустимых (различимых на выходе канала) цепочек любой длительности. Коротко говоря, канал определяет набор передаваемых цепочек, но не влияет на вероятности их передачи. Эти вероятности определяет источник, установленный на входе канала. Следовательно, производительность выхода канала H_i^* , рассматриваемого как составной источник, зависит не только от свойств самого канала, но и от свойств источника на его входе. Перебирая всевозможные источники на входе канала, можно выбрать такой, при котором H_i^* достигает максимума. Сам максимум, зависящий только от свойств канала, называется его *пропускной способностью* C (бит/с):

$$C = \max_{\substack{\text{по всем} \\ \text{источникам}}} H_i^*. \quad (2.14)$$

Руководствуясь свойствами энтропии, легко установить, что входной источник, обеспечивающий максимум H_i^* , должен назначать всем допустимым цепочкам канала достаточно большой длины равные вероятности появления. Таким образом, если число допустимых цепочек длины T составляет $N(T)$, то

$$C = \lim_{T \rightarrow \infty} \frac{\log N(T)}{T}. \quad (2.15)$$

Пусть теперь задан некоторый источник с производительностью H_i и канал связи без шумов с пропускной способностью C . Пусть между выходом источника и входом канала устанавливается некоторый преобразователь, призванный по возможности согласовать источник с каналом. Ответ на вопрос о предельных возможностях такого согласования дает следующая теорема («основная теорема для канала без шумов»).

Если $C - H_i > 0$ сколь угодно малая величина, то всегда существует такой преобразователь, который обеспечивает передачу информации источника по каналу без потерь. Передача информации со скоростью больше C невозможна, так что, если производительность источника $H_i > C$, то потери информации в канале будут не менее $H_i - C$, какой бы преобразователь между источником и каналом не использовался.

Для того чтобы не усложнять изложение, мы в данном разделе не будем приводить строгие доказательства формулируемых теорем. Однако пояснить идеи, лежащие в их основе, необходимо, поскольку они прямо указывают на перспективные пути совершенствования фундаментальных характеристик систем связи. Приведенная выше теорема, как и ряд других изложенных ниже положений теории информации, основываются на важнейших структурных особенностях выходных последовательностей источников. Сформулируем эти особенности применительно к последовательностям, описываемым цепями Маркова и рассматриваемым как последовательности состояний (см. 2.1.3).

Проанализируем некоторую последовательность на выходе источника длиной N состояний. С вероятностью, стремящейся к единице, по мере роста N в этой последовательности j -е состояние встретится около $V_j N$ раз, а переход (за один такт) из j -го состояния в i -е — около $V_j W_{j,i} N$ раз. Интерес представляет значение логарифма вероятности Q_k появления той

или иной (k -й) достаточно длинной последовательности (цепочки), так что саму эту вероятность достаточно вычислить с точностью до постоянного множителя. С требуемой точностью, в соответствии с вышеизложенным, для типовых цепочек, суммарная вероятность появления которых приближается к единице (с ростом N), имеем

$$Q_k = \prod_{j,i} W_{j,i}^{N V_j W_{j,i}}, \quad (2.16)$$

откуда

$$\frac{\log Q_k}{N} = \sum_{j,i} V_j W_{j,i} \log W_{j,i} = -H_f. \quad (2.17)$$

Обозначая число различных типовых цепочек через $M(N)$, найдем

$$\log M(N) = \log \frac{1}{Q_k} = N H_f. \quad (2.18)$$

Итак, все возможные цепочки источника достаточно большой длины распадаются на две группы: группа типовых цепочек с суммарной вероятностью, приближающейся с ростом N к 1, содержащая порядка $2^{N H_f}$ цепочек, и группа остальных, нетипичных цепочек с суммарной вероятностью, стремящейся к нулю. Важно отметить, что число типовых цепочек определяется энтропией источника, и фактически именно это обстоятельство в первую очередь обосновывает целесообразность принятого в теории информации определения данной величины.

Указанные свойства цепочек являются прямым следствием закона больших чисел, т.е. восходят к фундаментальным законам природы. Для теории связи выявленный факт означает, что при передаче информации, если система оперирует с достаточно длинными последовательностями, нужно, прежде всего, заботиться о передаче типовых цепочек информационных символов; нетиповые цепочки могут передаваться либо гораздо более протяженными сигналами, либо вообще не передаваться. Это мало повлияет на количество транслируемой информации. С учетом данного обстоятельства мы ниже вообще не будем интересоваться способами передачи нетиповых цепочек.

После изложенных замечаний легко пояснить суть сформулированной выше основной теоремы для канала без шумов. Действительно, легко установить, что если $H_i < C$, то число достаточно протяженных типовых цепочек источника не превысит числа цепочек той же длины, которые могут быть переданы по каналу. Кодирова (преобразуя) первые во вторые, можно обеспечить передачу информации без потерь. Если же $H_i > C$, то цепочек канала не хватит для кодирования типовых цепочек источника, и соответствующие потери информации окажутся неизбежными.

Ценность сформулированной теоремы для техники связи определяется следующим: если преобразование для согласования источника с каналом не производится, то для передачи сообщений требуется канал с пропускной способностью, равной объему сообщения V . Если же реализуется оптимальное согласование, то пропускная способность канала может быть сокращена до величины, как показывает теорема, H_i . Отсюда, чем ближе H_i к C , тем большая длина цепочек должна использоваться при кодировании. Увеличение длины кодовых комбинаций приводит к соответствующей задержке в передаче информации, так что за улучшение параметров системы связи приходится расплачиваться увеличением задержки. Это положение является общим для всех вскрытых теорией информации возможностей совершенствования систем телекоммуникаций.

2.3. Информация при недетерминированных преобразованиях сигналов. Пропускная способность канала с шумами

2.3.1. Информация при недетерминированных преобразованиях

Выше были рассмотрены преобразователи, для которых выходная последовательность есть заранее установленная функция входной. Расширим теперь рассматриваемые преобразования, вводя стохастические преобразователи, для которых связь между входными и выходными последовательностями является недетерминированной и задается статистически. Наиболее важным примером такого преобразователя служит канал с шумом, в котором к сигналу аддитивно добавляются шумы. Наша ближайшая задача — оценка количества информации, прошедшей через стохастический преобразователь.

Рассмотрим базовую схему, содержащую источник информации, выходной сигнал которого обозначим через x , и стохастический преобразователь, на выходе которого появляется последовательность цифр y , связанная с x статистически. Пусть энтропия источника H_x , а энтропия выхода преобразователя H_y . Введем также условные энтропии $H_{y/x}$ (величины y при известном x) и $H_{x/y}$ (величины x при известном y). Отметим, что в данном случае H_y может и превосходить H_x , поскольку теперь выход преобразователя содержит информацию не только о выборе, сделанном источником, но и о выборе, сделанном преобразователем. Если, например, источник производит всегда один и тот же выбор, т.е. его энтропия равна нулю, то y будет оставаться случайным, поскольку используется случайное преобразование; таким образом получим $H_y > 0$.

Рассмотрим еще энтропию совместной схемы, включающей пару испытаний x и y :

$$H_{x,y} = H_x + H_{y,x} = H_y + H_{x/y},$$

откуда, вводя новое обозначение $I_{x,y}$, запишем

$$I_{x,y} = H_x - H_{x/y} = H_y - H_{y/x} = I_{y,x}. \quad (2.19)$$

Величина $I_{x,y}$ представляет собой количество информации об x , которую можно получить, наблюдая y . Действительно, в соответствии с (2.19) она представляет собой неопределенность относительно x , которая предшествовала наблюдению y за вычетом той неопределенности, которая осталась после такого наблюдения. Таким образом, она равна неопределенности относительно x , снятой наблюдением y , т.е. определяет количество информации об x , содержащейся в y . Выражение (2.19) показывает, что количество информации в y об x , такое же, как и в x об y . По этой причине $I_{x,y}$ носит название *взаимной информации* между x и y . Учитывая (2.5), найдем, что взаимная информация всегда неотрицательна и равна нулю тогда и только тогда, когда x и y независимы.

Информационный баланс в рассматриваемой базовой схеме испытаний может быть условно описан конструкцией, представленной на рис. 2.3. Здесь показана информация H_x об x , поступающая на вход статистического преобразователя. Часть этой информации ($H_{x/y}$)

теряется, так что об x остается только информация $I_{x,y}$. Одновременно преобразователь вводит дополнительную информацию $H_{y/x}$, но уже о собственных статистических свойствах. В результате и формируется выходная информация преобразователя H_y .

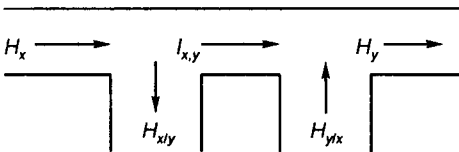


Рис. 2.3. Информационный баланс при стохастических преобразованиях

Как уже отмечалось, $I_{x,y}$ рассматривается в теории информации как количество информа-

ции об x в y . Кроме сформулированных выше свойств этой величины, разумность такого определения подтверждается еще следующим примером. Пусть, кроме базовой схемы, включающий источник и статистический преобразователь, существует еще некий «демон», которому доступно неискаженное наблюдение выходов как источника, так и преобразователя. Пусть этот демон передает наблюдателю выхода преобразователя дополнительные корректирующие данные, позволяющие полностью восстановить всю информацию об x , потерянную при преобразовании. Легко подсчитать, что требуемое количество дополнительной информации составит именно $H_{x/y}$, так что есть все основания рассматривать эту величину как потерю информации об источнике при преобразовании. Докажем данное утверждение для простейшего случая. Пусть речь идет о двоичном источнике и преобразователе, правильно воспроизводящем цифру источника с вероятностью p и заменяющем ее на обратную с вероятностью $q = 1 - p$. Дополнительная информация должна предусматривать две возможности: цифра, поступившая с выхода преобразователя, верна (вероятность такого исхода p) и цифра не верна (вероятность исхода q). Соответствующее количество дополнительной информации — $p \log p - q \log q$ как раз и есть $H_{x/y}$ в данном случае.

2.3.2. Канал связи с шумами и основная теорема для него

Канал связи с шумами полностью эквивалентен стохастическому преобразователю. Поэтому описанная в предыдущем пункте базовая схема может быть использована для оценки объемов информации, транслируемой таким каналом. При этом в качестве x должны рассматриваться цепочки источника, а в качестве y цепочки на выходе канала. Количество информации источника, передаваемой по каналу, может нормироваться либо к единице времени (I_t) либо к символу выходных цепочек (I_f). Если взаимную информацию для схемы, оперирующей с цепочками длины T обозначить через $I_{x,y}(T)$, а с цепочками с числом цифр по выходу канала N через $I_{x,y}(N)$, то можно записать

$$I_t = \lim_{T \rightarrow \infty} \frac{I_{x,y}(T)}{T}; \quad I_f = \lim_{T \rightarrow \infty} \frac{I_{x,y}(N)}{N}. \quad (2.20)$$

Отметим, что количество информации, передаваемой через канал, зависит не только от свойств канала, но и от характеристик установленного на его входе источника. Варьируя источники, можно изменять это количество. В частности, можно отыскать такой источник, при котором указанная величина достигает максимума. Значение максимума, зависящего уже только от свойств канала, называется *пропускной способностью канала связи с шумами*:

$$C = \max_{\text{по всевозможным источникам}} I_t. \quad (2.21)$$

Сформулируем теперь основную теорему теории информации для канала с шумами.

Если информация источника с производительностью R подлежит передаче по каналу с пропускной способностью C , причем $C > R$, то может быть найден такой детерминированный преобразователь (система кодирования) сообщений источника, что частота ошибок в цифрах переданного сообщения (потеря информации в канале) будет сколь угодно малой. Если же $C < R$, то передача с потерями информации меньшими $R - C$ невозможна.

Доказательство теоремы основывается на использовании свойств достаточно длинных цепочек, установленных в 2.2.2. В связи с этим ограничимся анализом возможностей передачи только типовых цепочек источника. Рассмотрим тот источник, который реализует пропускную способность канала. Обозначим его через S в отличие от фактически установленного на входе канала источника g .

Оперируя ниже с энтропиями для источника S , условимся рассматривать их значения за единицу времени. Для того, чтобы подчеркнуть это обстоятельство, введем в их обозначения дополнительный индекс t , записывая $H_{t,y}$, $H_{t,x/y}$ и $H_{t,y/x}$. Отметим, что здесь, как и выше, x — выход источника, y — выход канала.

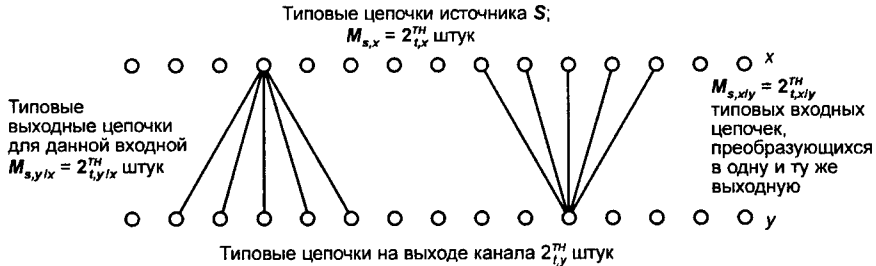


Рис. 2.4. Преобразования типовых цепочек в канале с шумом

На рис. 2.4 наглядно представлены основные соотношения между типовыми цепочками на входе и выходе канала связи для источника S . По определению $H_{t,x} - H_{t,x/y} = C$. Изображенные на рисунке два «веера» получаются благодаря использованию описанных выше свойств цепочек применительно к условным испытаниям, характеризуемым соответствующими условными энтропиями. Рассмотрим теперь типовые выходные цепочки источника g ($M_{g,x} = 2^{TR}$ штук) и закодируем их случайным образом в типовые выходные цепочки источника S . Вероятность того, что некоторая цепочка последнего множества (из x , рис. 2.4) окажется использованной как сообщение, составляет $M_{g,x}/M_{s,x}$. Ошибка при приеме этого сообщения не будет происходить, если ни одна другая цепочка правого веера (см. рис. 2.4) не будет использована как сообщение. Вероятность данного события при рассматриваемом случайном кодировании составляет

$$P = [1 - M_{g,x}/M_{s,x}] M_{s,x/y}. \quad (2.22)$$

Используем теперь известное соотношение, согласно которому в первом приближении $(1 + d)^h = 1 + dh$, если только $dh \ll 1$. Проверим, применимо ли оно к (2.22):

$$dh = M_{s,x/y} M_{g,x}/M_{s,x} = 2^T (R - H_{t,x} + H_{t,x/y}).$$

Но $H_{t,x} - H_{t,x/y} = C$. Пусть теперь $R < C$, так что можно записать $C - R = v > 0$. Тогда получим $dh = 2^{-vT}$, так что как ни было бы мало v , при достаточно большом T величина dh сколь угодно мала. По этой причине вероятность ошибки при описанном кодировании

$$P_{\text{ош}} = 1 - P = 2^{-vT} \rightarrow 0, \quad (2.23)$$

что и доказывает первое утверждение теоремы. Второе ее утверждение следует непосредственно из определения пропускной способности канала.

Основная теорема для канала с шумами имеет фундаментальное значение для техники связи, поскольку устанавливает потенциально достижимые значения важнейших характеристик систем передачи информации. Отметим, что эти достижимые значения оказываются существенно более высокими, чем достигнутые на практике, не только на момент появления основополагающей работы К. Шеннона (1948), но и к настоящему времени. Для иллюстрации такого положения достаточно указать на то выявленное теорией информации обстоятельство, что только за счет достаточно эффективного кодирования информации на входе канала связи при сохранении постоянной скорости передачи информации и мощно-

сти сигнала можно сделать ошибку передачи сообщений сколь угодно малой. Для всех реализованных систем стремление снизить вероятность ошибки всегда приводит к необходимости снижения скорости передачи (в пределе до нулевой) или увеличения мощности сигнала (в пределе неограниченного).

Хотя современная техника связи еще весьма далека от сколько-нибудь полного использования выявленных теорией возможностей, тем не менее на этом пути уже достигнуты впечатляющие результаты. Появление теории информации стимулировало бурное развитие теории кодирования, вызвавшее изобретение очень большого количества кодов, среди которых встречаются и весьма эффективные. Разработка новых, все более эффективных кодов продолжается и за последние десятилетия наблюдается устойчивый рост достигнутых показателей в этой области.

2.4. Предельная энергетическая и спектральная эффективность типовой системы связи

2.4.1. Пропускная способность типового канала связи

В качестве важнейшего примера применения результатов предыдущего параграфа оценим достижимые характеристики системы связи, использующей некий типовой канал связи. В качестве такового примем канал со строго ограниченной полосой пропускания (в области положительных частот, равной F) и аддитивным флуктуационным шумом. Спектральную плотность шума N_0 будем считать постоянной на всех частотах, его значения в любые сколь угодно близкие моменты времени независимыми, а их распределения вероятностей нормальными. Иначе говоря, речь идет об уже упоминавшейся в гл. 1 модели БГШ. Мощность шума в канале связи составляет $P_{\text{ш}} = N_0 F$. Что касается полезного сигнала, то примем, что, кроме уже отмеченного ограничения по полосе частот, ограничена еще и его средняя мощность, которая не должна превышать величину P_c . Примем также, что частотная характеристика канала в пределах его полосы пропускания есть константа. Такой канал весьма близок к каналам, используемым во многих реальных системах связи. Представляет интерес определение его пропускной способности.

Особенность сформулированной задачи заключается в том, что заданный в ней канал связи является аналоговым. Между тем все понятия и положения теории информации формулировались ранее применительно только к цифровым сигналам. По этой причине нам придется несколько развить принятый ранее подход.

Как отмечалось в предыдущей главе, любой радиосигнал с ограниченной полосой F однозначно определяется дискретными отсчетами своих видеозэквивалентов при суммарной частоте отсчетов $F_0 = 2F$. Пусть для определенности в качестве видеозэквивалентов сигналов используются их квадратурные проекции. Каждая из них представляет собой видеосигнал со строго ограниченной частотой $F_b = F/2$ энергетическим спектром. На выходе канала к отсчетам квадратур полезного сигнала x прибавляются отсчеты шума z , так что наблюдатель получает величину $y = x + z$ и выделяет на этой основе информацию $I_{x,y}$. Плотность p_z распределения вероятностей z для каждого отсчета нормальная с нулевым средним значением и дисперсией $D_z = N_0 F_b = 0,5 P_{\text{ш}}$. Различные отсчеты шума независимы.

Для определения пропускной способности канала статистические характеристики отсчетов полезного сигнала должны варьироваться до достижения максимального значения количества переданной по каналу информации. В соответствии со свойствами энтропии можно сразу сказать, что, прежде всего, для достижения искомого максимума отдельные отсчеты

полезного сигнала должны быть приняты независимыми. Поэтому дело сводится к максимизации количества информации, переносимого каждым отсчетом в отдельности. Это может быть достигнуто рациональным выбором плотности вероятностей отсчета полезного сигнала p_x . Для нее задана только дисперсия $D_x = 0,5 P_c$. Следовательно, $D_y = 0,5(P_c + P_{\text{ш}})$.

Для того чтобы применить к непрерывным величинам соотношения, полученные нами выше для цифровых величин, проведем квантование величин x , y и z с достаточно малым шагом δ . Тогда получим

$$H_x = - \int_{-\infty}^{\infty} \{p_x(u) \log [\delta p_x(u)]\} du = - \int_{-\infty}^{\infty} p_x(u) \log p_x(u) du - \log \delta. \quad (2.24)$$

Это выражение показывает, что энтропия аналоговой случайной величины стремится к бесконечности по мере роста точности ее наблюдения ($\delta \rightarrow 0$). Это соответствует общим положениям теории информации, поскольку означает неограниченный рост числа возможных исходов статистического испытания. В нашем случае величина x наблюдается с вполне конечной точностью из-за наличия шумов канала, что и приведет, как мы увидим ниже, к конечному значению объема транслируемой информации.

Далее $H_{y/x}$ определяется только распределением величины z и по аналогии с (2.24) может быть записано в виде

$$H_{x/y} = - \int_{-\infty}^{\infty} p_z(u) \log p_z(u) du - \log \delta. \quad (2.25)$$

Откуда

$$I_{x,y} = H_y - H_{y/x} = - \int_{-\infty}^{\infty} p_y(u) \log p_y(u) du + \int_{-\infty}^{\infty} p_z(u) \log p_z(u) du. \quad (2.26)$$

Важно отметить, что эта величина является уже конечной и не зависит от δ , так что из-за наличия шумов количество информации, передаваемой даже с помощью аналоговых сигналов за ограниченное время, остается конечным.

Величину (2.26) необходимо максимизировать, варьируя распределение p_x . Можно показать, что максимум достигается при нормальном распределении x , когда y имеет, соответственно, также нормальное распределение (сумма двух нормальных величин). Последний интеграл в (2.24), так же как и интеграл в (2.25), носит название *дифференциальной энтропии* H^d для соответствующих распределений (p_y и p_z соответственно). Дифференциальная энтропия нормальной случайной величины легко вычисляется:

$$H^d = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \left[\exp\left(-\frac{u^2}{2\sigma^2}\right) \right] \log \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{u^2}{2\sigma^2}\right) \right] du = 0,5 \log(2\pi\sigma^2) + \log e,$$

где σ — среднеквадратичное значение случайной величины.

Используя это соотношение для вычисления дифференциальных энтропий, входящих в (2.26) и учитывая, что, как отмечалось, нормальное распределение сигнала на выходе источника максимизирует объем передаваемой информации, находим максимальное значение количества информации, переносимой одним отсчетом:

$$I_0 = 0,5 \log [(P_c + P_{\text{ш}})/P_{\text{ш}}],$$

так что

$$C = 2FI_0 = F \log \frac{P_c + P_{\text{ш}}}{P_{\text{ш}}}. \quad (2.27)$$

Последняя формула и есть знаменитая формула Шеннона, содержание которой анализируется ниже.

2.4.2. Достижимые показатели эффективности систем связи

Всякая система связи может быть охарактеризована двумя основными показателями, определяющими эффективность использования ею энергетических и полосных ресурсов. Под *энергетическим ресурсом* понимается здесь ВЧ-мощность радиосигналов, а под *полосным* — занимаемая им полоса частот. Энергетическая эффективность системы связи определяется нормированным отношением сигнал/шум, которое должно быть создано на приемной стороне радиолинии для обеспечения заданного качества связи. Применительно к интересующей нас цифровой связи, это означает обеспечение работы с вероятностью неправильного воспроизведения символа сообщения $P_{\text{ош}}$, не превосходящей заданную. Упомянутое выше нормированное отношение сигнал/шум измеряется как отношение на входе приемника (выходе приемной антенны) энергии сигнала E_6 , расходуемой на передачу одного бита сообщения, к спектральной плотности N_0 всех шумов радиолинии, пересчитанных к той же точке

$$h_6^2 = E_6 / N_0. \quad (2.28)$$

Величина определенного таким образом отношения, необходимого для обеспечения требуемого значения $P_{\text{ош}}$, является универсальным показателем энергетической эффективности того или иного метода передачи информации. Независимо от вида используемых сигналов, способов модуляции и демодуляции, кодирования и декодирования, лучшим с точки зрения использования энергетического потенциала является тот вариант, для которого требуется меньшее значение h_6 , ибо это означает, что при одинаковых шумах канала (по спектральной плотности) в этой радиолинии расходуется меньшая энергия передатчика для передачи каждого бита сообщения. Отметим, что если средняя скорость передачи информации составляет I_i , то среднее время передачи одного бита $T_6 = 1/I_i$, и $E_6 = P_c T_6 = P_c / I_i$. Полоса частот F_n численно равна I_i , именуется *информационной полосой*, так что h_6 можно рассматривать как отношение мощности сигнала к мощности шумов $P_{\text{ш}}$ в информационной полосе:

$$h_6^2 = P_c / (N_0 F_n) = P_c / P_{\text{ш}}. \quad (2.29)$$

Что касается эффективности использования системой передачи информации полосы рабочих частот, то она характеризуется отношением реализованной в данной системе скоростью передачи информации к ширине полосы занимаемых рабочих частот

$$\gamma = I_i / F. \quad (2.30)$$

Эта величина именуется *удельной скоростью передачи информации* и измеряется в $\frac{\text{бит/с}}{\text{Гц}}$.

Перепишем теперь формулу Шеннона (2.27), используя введенные показатели эффективности. При этом учтем, что скорость передачи информации в оптимальной системе, фигурирующей в формуле Шеннона, $I_i = C$. Итак, $C = F \log[1 + P_c / N_0 F]$, или $\log[1 + P_c C / N_0 F C] = C / F = \gamma$. Иначе, $\gamma = \log(1 + \gamma h_6^2)$, или

$$h_6^2 = \frac{2^\gamma - 1}{\gamma}. \quad (2.31)$$

Это выражение устанавливает связь между двумя главными показателями канала связи, характеристиками его энергетической и полосной (спектральной) эффективности для случая наилучшего возможного использования этого канала. Последнее утверждение следует из того обстоятельства, что (2.31) получено в предположении применения методов передачи, полностью реализующих пропускную способность рассматриваемого канала связи.

Построенная в соответствии (2.31) зависимость между h_6 и γ представлена на рис. 2.5. Точки заштрихованной области реализуемы, а лежащие вне ее реализованы быть не могут. Наиболее эффективные системы связи соответствуют точкам заштрихованной области, лежащим вблизи ее границы. Эта граница является общей для всех систем независимо от допустимой вероятности ошибки воспроизведения информации в этих системах, что следует из основной теоремы для канала с шумами.

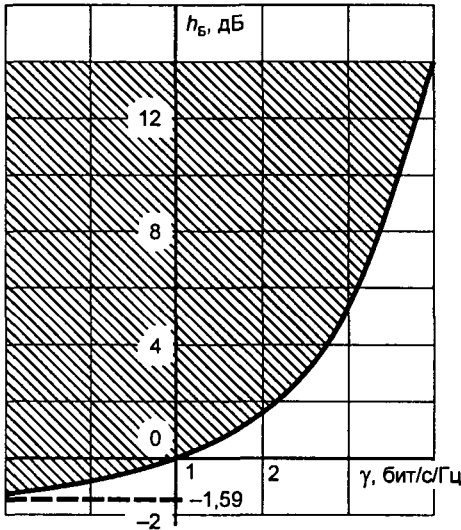


Рис. 2.5. Граница достижимых значений эффективности систем передачи информации

При γ стремящемся к нулю, как следует из (2.31),

$$h_6 \rightarrow -1,59 \text{ дБ.} \quad (2.32)$$

Соответствующая асимптота проведена на рис. 2.5 штриховой линией.

Для больших значений γ справедливо соотношение

$$h_6 [\text{дБ}] = 3,01\gamma - 10 \log \gamma. \quad (2.33)$$

Интересно посмотреть, где на данной диаграмме лежат точки, соответствующие современным системам связи. Например, для системы сотовой связи стандарта GSM в отсутствие затенений и замираний на трассе распространения реализуются следующие показатели: $h_6 = 5$ дБ (вероятность ошибки 10^{-5}) и $\gamma = 0,675$ бит/с/Гц. Видно, что даже в этой весьма совершенной системе имеется

неиспользованный резерв по энергетической эффективности, составляющий около 6 дБ. При полном использовании данного резерва можно было бы увеличить γ до 4 бит/с/Гц, т.е. увеличить ее пропускную способность почти в шесть раз, не меняя ни мощность передатчиков, ни чувствительность приемников, ни используемую полосу частот. Освоение всех возможностей требует разработки новых, более совершенных систем кодирования, для чего необходимо достаточно много времени. Внедрение современных турбокодов позволит использовать еще 1,5–2 дБ из указанного выше резерва энергетической эффективности.

Известен один гипотетический вариант передачи информации, параметры которого лежат на представленной на рис. 2.5 границе. Речь идет о системе с ортогональными, нормированными по мощности сигналами при неограниченно возрастающей кратности кода k (значность кода $m = 2^k$). Если передаче подлежит двоичное сообщение, то вначале последовательные цифры сообщения собираются в данной системе в блоки по k цифр в каждом. Каждому блоку ставится в соответствие один из ортогональных сигналов равных энергий. Осуществляется идеальный когерентный прием этих сигналов. Анализ показывает (см. § 4.3), что для такого ортогонального кодирования h_6 стремится к 1,59 дБ при γ , стремящемся к нулю, что в точности соответствует предельным возможностям, устанавливаемым теорией информации. К сожалению, такой случай соответствует предельно низкой полосной эффективности.

Рассмотрим еще один реальный пример, характеризующийся сравнительно высокой полосной эффективностью. Речь идет о проводных модемах для передачи компьютерной информации по телефонным каналам тональной частоты (ТЧ). Современные устройства обеспечивают передачу со скоростью до 56 кбит/с. Так как полоса пропускания ТЧ-канала (от 0,3 до 3,4 кГц) составляет 3,1 кГц получим $\gamma = 18$ бит/с/Гц. В соответствии с (2.31) это требует $h_6 = 32,6$ дБ. По действующим нормам соотношение сигнал/шум в ТЧ-канале должно быть не менее величин порядка 50 дБ, так что формально требуемое значение должно обеспечиваться с необходимым запасом. На случай, если реально оно не обеспечивается, в проводных модемах предусматривают возможность передачи и с меньшими скоростями (начиная с 2,4 кбит/с).

Анализируя общее поведение кривой рис. 2.5, отметим, что оно существенно меняется в районе значений γ порядка 1. Даже за умеренное увеличение полосной эффективности сверх этого значения приходится расплачиваться резким снижением энергетической эффективности; наоборот, ниже этого значения увеличение удельной скорости передачи информации достигается за счет сравнительно небольшого увеличения h_6 .

ИДЕАЛЬНЫЙ ПРИЕМ ЦИФРОВЫХ СИГНАЛОВ

3.1. Постановка задачи идеального приема при цифровой передаче сообщений

3.1.1. Структура рассматриваемых радиолиний

Структура радиолинии передачи цифровой информации, которую мы будем рассматривать в настоящей главе, представлена на рис. 3.1. Она является в известном смысле простейшей, поскольку дополняется еще рядом устройств, обеспечивающих улучшение тех или иных характеристик радиолинии. Такие устройства будут вводиться в структуру по мере изучения в последующих разделах. Ограничимся только устройствами, показанными на рис. 3.1.

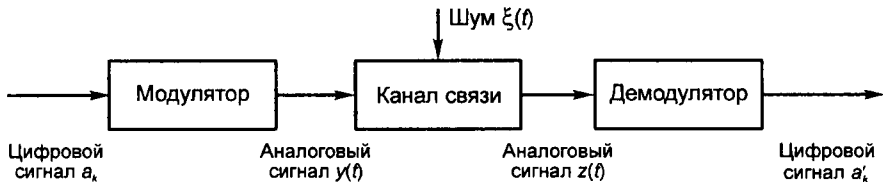


Рис. 3.1. Структура рассматриваемой радиолинии

Под цифровым сигналом здесь понимаются в общем случае последовательности m -ичных цифр, следующих во времени одна за другой со скоростью B цифр в секунду. Таким образом, среднее время, расходуемое на передачу одной двоичной цифры, составляет

$$T_6 = \frac{\log_2 m}{B}, \quad (3.1).$$

а одной m -ичной

$$T_3 = 1/B. \quad (3.1 \text{ а})$$

Модулятор и демодулятор на рис. 3.1 понимаются в широком смысле как комплекс устройств, обеспечивающих согласование аналогового канала связи с транслируемым цифровым сигналом. Таким образом, модулятор, кроме собственно модулятора, включает и преобразователи частоты, и усилители, в том числе и передатчик, и передающую антенну. Аналогично и демодулятор включает приемную антенну, усилители с преобразованиями частоты и, конечно, собственно демодулятор.

Что касается канала, то здесь предполагается, что он не искажает формируемые модулятором сигналы, а лишь изменяет их абсолютный уровень в соответствии со своим коэффи-

циентом передачи. Для того, чтобы упростить изложение, ниже коэффициент передачи канала связи принимается равным единице, а его истинное значение учитывается в уровне выходного сигнала модулятора $y(t)$.

Принцип работы модулятора заключается в том, что при поступлении на его вход очередной, например j -й m -ичной цифры a_j , на его выходе формируется соответствующий ей сигнал $s_j(t)$. Поскольку a_j может принять m значений, то и для сигнала $s_j(t)$ должно существовать, по крайней мере, столько же форм, каждая из которых именуется *элементарным сигналом* (ЭС). В общем случае, однако, по причинам указанным ниже, число ЭС может превышать m . При этом каждому значению входной цифры модулятор ставит в соответствие один из нескольких или даже целого ряда таких сигналов, причем выбор действительно используемого ЭС каждый раз происходит случайно. Таким образом, ансамбль всех используемых ЭС делится на m непересекающихся подмножеств, каждое из которых ставится в соответствие своему значению транслируемой цифры. Чаще всего различные ЭС отличаются значениями тех или иных параметров. Поэтому будем записывать сигналы в виде $S_j = S(\mu, \chi, t)$, где $j = 1, 2, \dots, m$ — значение транслируемой цифры, так что χ — манипуляционный или информационный параметр, а ряд значений параметра μ (обычно векторного) при данном j определяет набор ЭС, входящих в j -е подмножество. Параметры μ будем именовать *вспомогательными*. Появление этих параметров определяется тем фактом, что реально невозможно абсолютно точно формировать заданные сигналы, тем более, если как в данном случае речь идет о формировании на приемной стороне радиолинии образцов сигналов, используемых модулятором на ее передающей стороне.

Что касается временной структуры сигналов, то принимается, что в последовательности передаваемых по радиолинии ЭС каждый очередной сигнал занимает тактовый интервал длиной T_s . В частности, на выходе канала границы тактовых интервалов определяются тактовыми точками, моменты появления которых $t_k = t_0 + kT_s$, $k = \dots, -1, 0, 1, \dots$. Именно в тактовых точках может происходить смена передаваемого ЭС.

В соответствии с типовой реальной ситуацией ниже принимается, что в канале связи на сигнал $y(t)$ накладываются шумы, так что на выходе канала появляется напряжение

$$z(t) = y(t) + \xi(t). \quad (3.2)$$

Демодулятор радиолинии должен определять номер j подмножества, которому принадлежит принимаемый на каждом тактовом интервале сигнал. Ясно, что шумы канала делают этот процесс отнюдь не тривиальным. Более того, как правило, реализация шума может принять с известной вероятностью любую форму, так что абсолютно надежное определение j вообще не возможно; какой бы алгоритм решения данной задачи не использовался, всегда остается определенная вероятность $P_{\text{ош}}$ ошибочного решения. Именно поиск алгоритма работы демодулятора, обеспечивающего минимальное значение этой вероятности, и является одной из основных задач, рассматриваемых в настоящем разделе.

Анализ путей построения реальных демодуляторов показывает, что все они включают устройства двух типов:

- собственно демодулятор;
- вспомогательные устройства, определяющие вспомогательные параметры μ ЭС.

Дело в том, что для демодуляции всегда необходимо знание в том или ином объеме вспомогательных параметров ЭС, таких как положение тактовых точек принимаемого сигнала, определяемое параметром t_0 , истинное значение несущей частоты принимаемых ЭС, иногда — фазы несущей частоты и уровня сигнала на выходе канала и др. Все перечисленные параметры изменяются достаточно медленно, так что их можно считать постоянными в течение ряда следующих один за другим ЭС. Поэтому, как правило, работа вспомогатель-

ных систем основывается на анализе $z(t)$ на достаточно протяженных временных интервалах, во много раз превосходящих T_s .

В отличие от этого, собственно демодулятор основывает свои решения только на значениях $z(t)$ на данном тактовом интервале, обновляя входные данные и принимаемые решения каждый тактовый интервал. При этом он использует данные о вспомогательных параметрах ЭС, вырабатываемые вспомогательными системами.

3.1.2. Постановка задачи идеального приема

Теория идеального приема была развита нашим соотечественником В.А. Котельниковым в середине XX в. Она послужила результатом поиска алгоритмов работы демодуляторов, обеспечивающих наилучшее возможное качество их работы, т.е. применительно к цифровым демодуляторам минимально возможное значение вероятности ошибки $P_{\text{ошз}}$ при заданном уровне сигнала и шумов. В то время уже сам факт существования таких алгоритмов вызывал сомнения. Успешное решение поставленной задачи явилось следствием ее удачной постановки. Котельников отказался, по существу, от поиска оптимальных алгоритмов работы вспомогательных систем, сосредоточив внимание на поиске алгоритмов работы собственно демодулятора. Это было достигнуто с помощью предположения, что все вспомогательные параметры ЭС, используемые демодулятором, заранее известны на приемной стороне радиолинии. Такое предположение позволило получить важные, содержательные результаты в области идеального приема и обойти проблемы оптимального построения вспомогательных систем, многие из которых не решены и по сию пору.

Подчеркнем теперь, что сформулированное предположение не означает, что все вспомогательные параметры ЭС в теории идеального приема всегда полагаются известными на приемной стороне радиолинии. Хотя определенный раздел теории действительно исходит из такого предположения (теория когерентного приема сигналов), рассматривается и случай, когда некоторые вспомогательные параметры принимают случайные значения. Полагается, однако, что на каждом тактовом интервале вспомогательные параметры принимают новые случайные значения, не зависящие от их значений на других тактовых интервалах. В этих условиях вспомогательные параметры могут при необходимости определяться только путем анализа отрезка сигнала, принятого на данном тактовом интервале, что относится к компетенции уже собственно демодулятора, а не каких-либо вспомогательных систем.

Сформулируем еще одно важное предположение, лежащее в основе излагаемой ниже теории идеального приема сигналов. Будем считать, что ЭС, передаваемые на различных тактовых интервалах, не имеют между собой какой-либо связи. Такая связь может возникать как за счет взаимозависимости между отдельными цифрами сообщения, поступающими на вход модулятора, так и из-за взаимосвязи между вспомогательными параметрами ЭС, формируемых модулятором на различных тактовых интервалах. Очевидно, что если такая связь существует, то идеальная демодуляция сигнала на данном тактовом интервале не может проводиться независимо от демодуляции связанных с ним сигналов других тактовых интервалов. По этой причине, если не принять предположения о независимости сигналов на различных тактовых интервалах, то задача поиска алгоритма идеального приема чрезмерно усложнится. Кроме того, как показала теория информации (см. гл. 2), именно установление рациональных взаимосвязей между последовательно передаваемыми ЭС является основным направлением совершенствования одной из ключевых характеристик любой системы передачи информации — ее помехоустойчивости. Поэтому и полностью отказываться в теории

идеального приема от учета взаимозависимости последовательно передаваемых ЭС нецелесообразно. Из этого противоречия может быть указан следующий выход. В теории идеального приема всегда будем считать элементарные сигналы, передаваемые на различных тактовых интервалах, независимыми. Но вот само понятие «элементарный сигнал» будем трактовать по-разному в зависимости от того, имеется ли взаимосвязь между значениями $y(t)$ на различных тактовых интервалах. Такая зависимость, если и имеет место, то существенно убывает по мере увеличения расстояния (по времени) между рассматриваемыми тактовыми интервалами. Поэтому, если рассмотреть две последовательные цепочки ЭС (например, цепочку, занимающую тактовые интервалы с номерами от $-2n$ до 0 и от 1 до $2n$), то верно утверждение о том, что с ростом n можно обеспечить сколь угодно слабую взаимосвязь между ЭС, передаваемых на средних тактовых интервалах таких цепочек (с номерами около $-n$ и n). В связи с этим для идеальной демодуляции при наличии взаимосвязи между сигналами, следует перейти к рассмотрению укрупненных ЭС, в качестве которых рассматривать все те же цепочки длиной $2n$, которые могут иметь место с учетом существующих связей между символами. Теперь для идеальной демодуляции сигнала, занимающего, например, l -й тактовый интервал, необходимо строить идеальный приемник, предназначенный для идеального приема укрупненных ЭС, занимающих тактовые интервалы с номерами от $l-n$ до $n+l$, считая их независимыми от предыдущих и последующих укрупненных ЭС. По укрупненному сигналу, полученному в результате такой демодуляции, и должен определяться номер ЭС, переданного на l -м тактовом интервале. Проведенная таким образом демодуляция и будет в первом приближении идеальной.

Таким образом, принимаемое в теории идеального приема предположение о независимости элементарных сигналов не исключает использование полученных результатов в тех случаях, когда такая зависимость фактически существует.

В заключение настоящего параграфа остановимся еще на предположениях относительно характеристик шумов канала связи, принимаемых в теории идеального приема. Эта теория ориентирована, прежде всего, на флуктуационные шумы, причем, по причинам, о которых говорилось выше, в качестве модели шумов в теории идеального приема рассматривается белый гауссовский шум с постоянной спектральной плотностью, обозначенной ниже как N_0 .

3.2. Когерентный идеальный прием

3.2.1. Структура идеального приемника

При когерентном приеме все вспомогательные параметры ЭС предполагаются в точности известными на приемной стороне, так что теряет смысл сохранение параметра μ в обозначении сигналов. Применительно к данному случаю будем говорить, что по рассматриваемой радиолинии на k -м тактовом интервале, т.е. на отрезке времени от $t_k = t_0 + kT_s$ до $t_{k+1} = t_0 + (k+1)T_s$, передается ЭС $s_k(t)$, который может принять одно из m возможных значений: $S_j(t) = S(\mu, \chi, t)$; $j = 1, 2, \dots, m$ в зависимости от значения очередного транслируемого числа a_k .

Для наблюдателя, принимающего сигнал $z(t)$, выбор номера j , передаваемого ЭС, является случайным. Заранее (априори) предполагаются известными лишь вероятности P_j ($j = 1, 2, \dots, m$) выбора того или иного значения, подлежащего передаче числа, и, следовательно, того или другого ЭС. После анализа принятого на рассматриваемом тактовом интервале напряжения $z(t)$ у наблюдателя на приемной стороне появляются дополнительные сведения о выбранном в действительности сигнале. Объединением всех данных об этом событии

являются m апостериорных (послеопытных) вероятностей $P_n(j) = \mathfrak{Z}\{j/z(t)\}$, каждая из которых представляет собой вероятность того, что принятый отрезок $z(t)$ признается результатом наложения шумов на переданный j -й ЭС $S_j(t)$. При определении апостериорных вероятностей используются условные вероятности $\mathfrak{Z}\{z(t)/j\}$ того, что, если на рассматриваемом тактовом интервале был передан j -й ЭС, то на нем будет принят сигнал $z(t)$. Эта вероятность полностью совпадает с вероятностью того, что шум канала принял форму $\xi(t) = z(t) - S_j(t)$. Как было пояснено в 1.2, вероятность появления такой реализации шумов может быть оценена с помощью функционала плотности вероятности (1.55). Таким образом,

$$\mathfrak{Z}\{z(t)/j\} = W[z(t) - S_j(t)] = c \exp \left\{ -\frac{1}{N_0} \int_{t_k}^{t_{k+1}} [z(t) - S_j(t)]^2 dt \right\}. \quad (3.2)$$

Эту величину, рассматриваемую как функция от j , именуют *функцией правдоподобия*. Ее значения будем обозначать λ_j . Для определения апостериорных вероятностей рассмотрим совместные вероятности передачи j -го ЭС и появление на выходе канала на данном тактовом интервале сигнала $z(t)$:

$$\mathfrak{Z}\{j, z(t)\} = \mathfrak{Z}\{j\} \mathfrak{Z}\{z(t)/j\} = P_j \mathfrak{Z}\{z(t)/j\}.$$

Та же вероятность может быть представлена и в другом виде:

$$\mathfrak{Z}\{j, z(t)\} = \mathfrak{Z}\{z(t)\} \mathfrak{Z}\{j/z(t)\} = \mathfrak{Z}\{z(t)\} P_n(j).$$

Сопоставляя два последних соотношения, находим

$$P_n(j) = g P_j \mathfrak{Z}\{z(t)/j\} = g P_j \lambda_j, \quad (j = 1, 2, \dots, m), \quad (3.3)$$

где $g = 1/\mathfrak{Z}\{z(t)\}$ — коэффициент, не зависящий от j .

Знание апостериорных вероятностей позволяет решить задачу оптимального приема. Действительно, если принять в качестве номера переданного ЭС то значение j , для которого апостериорная вероятность (3.3) достигает максимума, то это и будет решение, соответствующее минимуму вероятности ошибки в приеме ЭС $P_{\text{ош}}$. Таким образом, алгоритм работы идеального приемника сводится к следующим процедурам:

- на каждом тактовом интервале для каждого j в соответствии с (3.2) вычисляются коэффициенты правдоподобия λ_j ; в соответствии с (3.3) — апостериорные вероятности;
- в качестве номера ЭС переданного на данном тактовом интервале принимается тот номер ЭС l , для которого достигается максимум апостериорной вероятности, т.е. такой, что $P_n(l) \geq P_n(j)$ для всех $j \neq l$.

В алгоритме при вычислении коэффициентов правдоподобия могут быть сделаны определенные упрощения. Во-первых, всегда можно считать, например, $g = 1$ и $c = 1$, поскольку данные коэффициенты не зависят от j и потому не влияют на результат работы по описанному алгоритму. В данном алгоритме апостериорные вероятности всегда могут вычисляться с точностью до постоянного, т.е. не зависящего от j множителя или слагаемого, поскольку это не повлияет на номер той вероятности, которая окажется максимальной.

Во-вторых, вместо априорных вероятностей можно в предлагаемом алгоритме рассматривать любую монотонно возрастающую функцию этих величин. В частности, вместо вычислений по формуле (3.3) можно принять с учетом (3.2):

$$P_n(j) = \ln P_j + \ln \lambda_j = N_0 \ln P_j + v_j, \quad (3.4)$$

где $v_j = - \int_{t_k}^{t_{k+1}} [z(t) - S_j(t)]^2 dt = - \int_{t_k}^{t_{k+1}} z^2 dt + 2 \int_{t_k}^{t_{k+1}} z S_j dt - \int_{t_k}^{t_{k+1}} S_j^2 dt$. При вычислениях v_j в соответст-

вии с последним соотношением можно опустить первый член, как не зависящий от j . Тогда, пренебрегая постоянным множителем, получим

$$P_n(j) = \frac{e_j}{2} + \rho_j, \quad (3.5)$$

где $\rho_j = \int_{t_k}^{t_{k+1}} z(t)S_j(t)dt$, $e_j = N_0 \ln P_j - E_j$ и $E_j = \int_{t_k}^{t_{k+1}} S_j^2(t)dt$ — энергия j -го ЭС.

В частном случае, когда априорные вероятности всех ЭС равны, выражение (3.4) можно упростить, сведя его к соотношению

$$P_n = \lambda_j = \rho_j - \frac{E_j}{2}. \quad (3.5a)$$

Изложенное показывает, что когда априорные вероятности различных ЭС различны, практическая реализация идеального прием заметно усложняются. Дело в том, что вычисление величин e_j требует знания абсолютного уровня сигнала и шумов на входе приемного устройства, для чего на практике необходимо вводить в состав приемника специальные измерительные устройства. В случае равенства априорных вероятностей знание уровня шума не требуется. Сверх того, в частном случае, когда энергии всех ЭС одинаковы (нормированные сигналы), не требуется и знание уровня полезного сигнала, поскольку член с E_j в (3.5a) перестает зависеть от j и может не учитываться.

С учетом априорных вероятностей связана одна принципиальная проблема. Дело в том, что вероятности появления на входе модулятора тех или иных значений чисел определяется абонентом, использующим данную радиолинию. Абонент не всегда знает эти априорные вероятности. Более того, вероятности в процессе использования радиолинии могут изменяться. Было бы весьма сложно соответственно изменять устройство демодулятора. Да и анализ показывает, что эффективность демодулятора, как правило, мало меняется при учете реальных изменений априорных вероятностей в достаточно широких пределах. По этой причине часто при построении демодуляторов принимают априорные вероятности всех сигналов равными. Алгоритм, основанный на таком предположении, называют *алгоритмом максимального правдоподобия*, в отличие от полной, сформулированной выше версии, именуемой *алгоритмом максимальной апостериорной вероятности*.

Найденные выше алгоритмы идеального приема могут быть представлены в виде соответствующих блок-схем идеальных демодуляторов. На рис. 3.2 представлена блок-схема идеального приема, реализующего алгоритм максимума правдоподобия. Приемник содержит m параллельно работающих устройств вычисления коэффициентов правдоподобия λ_j согласно (3.5a); на входы устройств параллельно подается сигнал $z(t)$, а их выходные напряжения поступают на устройство сравнения. Последнее определяет номер максимального коэффициента правдоподобия и выдает его на выход как результат демодуляции на данном тактовом интервале. Интеграторы в представленной схеме осуществляют интегрирование напряжения на своем входе в пределах данного тактового интервала от t_k до t_{k+1} . Приемник аналогичным образом демодулирует сигнал на следующем тактовом интервале и т.д. При смене обрабатываемого интервала соответственно сдвигаются во времени образцы ЭС и меняются пределы интегрирования в интеграторах. Основной частью вычислителей коэффициентов правдоподобия в этой схеме служат устройства, состоящие из генератора образца сигнала, перемножителя и интегратора. Такие устройства именуют *корреляторами*, поскольку их выходное напряжение ρ_j оценивает степень корреляции между принятым отрезком сигнала и данным образцом.

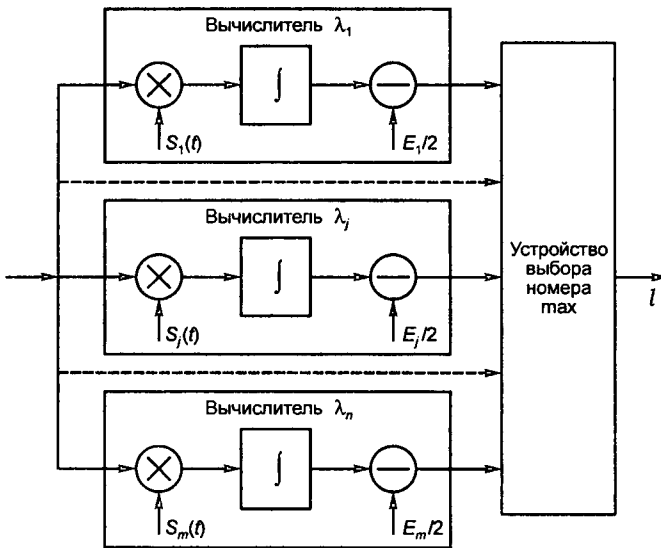


Рис. 3.2. Блок-схема идеального когерентного приемника

Отметим, что при приеме по максимуму апостериорной вероятности блок-схема идеального приемника мало изменяется; именно в блоках вычитание величин $E_j/2$ должно быть заменено на вычитание величин $e_j/2$. Однако, как было только что отмечено, это требует усложнения вспомогательных систем.

Естественно, что полученные выше математические выражения, описывающие работу идеального приемника, могут реализовываться различными техническими устройствами. В частности, в схеме на рис. 3.2 вместо корреляторов могут использоваться специальные фильтры, именуемые согласованными. Остановимся на их описании. Связь между входным $z(t)$ и выходным $y(t)$ сигналами всякого фильтра может быть описана интегралом Дюамеля:

$y(t) = \int_{-\infty}^{\infty} z(\tau)H(t-\tau)d\tau$, где $H(t)$ — импульсный отклик фильтра. Если положить, что

он отличен от нуля только на интервале от 0 до T_3 , то получим $y(t) = \int_{t-T_3}^t z(\tau)H(t-\tau)d\tau$. Для

того, чтобы фильтр был согласован с сигналом $S_j(t)$, его импульсный отклик должен иметь специальную форму:

$$H(t) = S_j(T_3 - t), \quad (3.6)$$

причем здесь принято, что сигналы $S_j(t)$ заданы так, что они отличаются от нуля только на интервале $(0, T_3)$. При таком выборе $H(t)$ получим

$$y(t_{k+1}) = \int_{t_k}^{t_{k+1}} z(t)S_j(t-t_k)dt, \quad (3.7)$$

что полностью совпадает с выходным напряжением коррелятора на k -м тактовом интервале. Таким образом, согласованные фильтры, если их отклики измерять в тактовых точках, могут заменить в идеальном приемнике корреляторы.

Отметим еще, что в ряде конкретных случаев схема идеального приемника может быть упрощена. Так, для случая нормированных сигналов, т.е. когда все ЭС имеют равные энергии, вычитание половины этой энергии во всех ветвях может быть опущено, поскольку оно не влияет на определение номера ЭС с максимальным коэффициентом правдоподобия. Иногда оказывается, что среди возможных ЭС имеются повторяющиеся друг друга по форме. Это имеет место, например, при двоичной фазовой телеграфии, когда ЭС отличаются только знаком. Такая же ситуация наблюдается при использовании биортого-

нальных ансамблей сигналов. Для такой пары сигналов может быть использована одна ветвь определения коэффициентов правдоподобия (один коррелятор), поскольку данные коэффициенты для таких сигналов жестко связаны. Так, для ФМ2 можно сохранить одну ветвь вычисления коэффициента правдоподобия первого ЭС λ_1 . Ясно, что коэффициент правдоподобия второго сигнала $\lambda_2 = -\lambda_1$. Поэтому, если $\lambda_1 > 0$, то $\lambda_1 > \lambda_2$, и наоборот, если $\lambda_1 < 0$, то $\lambda_1 < \lambda_2$. Таким образом, в идеальном приемнике ФМ2 можно опустить одну ветвь коррелятора и заменить устройство выбора номера сигнала, имеющего максимальное правдоподобие, простым пороговым устройством с нулевым порогом.

3.2.2. Потенциальная помехоустойчивость при двоичной передаче

Под *помехоустойчивостью* понимают способность приемника противостоять воздействию помех радиоприему в процессе демодуляции. Количественно помехоустойчивость характеризуется зависимостью вероятности ошибочной демодуляции от отношения уровня сигнала к уровню помех на входе демодулятора; чем меньшее значение этого отношения необходимо для приема сообщений с заданной достоверностью, тем выше помехоустойчивость приема. По определению самой высокой достижимой помехоустойчивостью обладает идеальный приемник. В связи с этим его помехоустойчивость именуется *потенциальной*.

Как показано в предыдущем параграфе, при полностью известных возможных сигналах идеальной является структура приемника, представленная на рис. 3.2. (При условии, что априорные вероятности всех сигналов равны, так что идеальным является прием по максимуму правдоподобия.) В соответствии с этим определением зависимости вероятности ошибки в приеме символа сообщения от отношения сигнал/шум на входе приемника, построенного по схеме рис. 3.2, и является главной задачей настоящего раздела. Отметим, что эту задачу удастся строго решить в достаточно общем виде только для случая двоичной передачи информации. Точные решения при недвоичной передаче известны только для некоторых частных случаев. В общем же случае могут быть получены отдельные оценки помехоустойчивости, из которых мы ниже приведем асимптотическую оценку, справедливую лишь для достаточно малых вероятностей ошибки.

Начнем со случая двоичной передачи. Пусть на приемной стороне в точности известны оба ЭС $S_1(t)$ и $S_2(t)$, которые могли быть переданы на данном тактовом интервале, в качестве границ которого без ограничения общности могут быть приняты моменты времени 0 и T_6 . (В данном случае $T_3 = T_6$, поскольку каждый элементарный сигнал соответствует передаче одного бита.)

Задачу вычисления помехоустойчивости будем решать следующим образом: примем, что в действительности был передан, например, первый сигнал и найдем вероятность того, что приемником будет зарегистрирован второй сигнал, что и определит вероятность ошибки. В соответствии с этим предположением сигнал, поступающий на вход приемника на рассматриваемом тактовом интервале, в действительности:

$$z(t) = S_1(t) + \xi(t), \quad (3.8)$$

где $\xi(t)$ — реализация шума БГШ канала, имевшая место на этом тактовом интервале.

Учитывая алгоритм работы идеального когерентного приемника, отображаемый выражением (3.5а), найдем, что вычисленные приемником коэффициенты составят:

$$\lambda_1 = \rho_1 - \frac{E_1}{2} = \int_0^{T_6} S_1^2(t) dt + \int_0^{T_6} \xi(t) S_1(t) dt - \frac{E_1}{2} = \frac{E_1}{2} + \zeta_1 \quad (3.9)$$

и

$$\lambda_2 = \rho_2 - \frac{E_2}{2} = \int_0^{T_6} S_1(t)S_2(t)dt + \int_0^{T_6} \xi(t)S_2(t)dt - \frac{E_2}{2} = \sqrt{E_1 E_2} r - \frac{E_2}{2} + \zeta_2. \quad (3.10)$$

В этих выражениях обозначено $\zeta_j = \int_0^{T_6} \xi(t)S_j(t)dt$ и $r = \frac{1}{\sqrt{E_1 E_2}} \int_0^{T_6} S_1(t)S_2(t)dt$.

В соответствии с известным из математики неравенством Коши величина r по модулю не превосходит единицу. Ее называют *нормированным коэффициентом корреляции сигналов*. Легко видеть, что $r = 1$ тогда и только тогда, когда $S_1(t) = KS_2(t)$ и $r = -1$, когда эти сигналы противоположны: $S_1(t) = -KS_2(t)$, где K — любая положительная константа. Если $r = 0$, то сигналы называются *ортогональными*.

Выражения (3.9) и (3.10) показывают, что коэффициенты правдоподобия принимают случайные значения, поскольку в них входят случайные величины ζ_1 и ζ_2 , зависящие от конкретной реализации шума на рассматриваемом тактовом интервале. В согласии с алгоритмом работы идеального приемника ошибка в демодулированном символе возникает, если оказывается $\lambda_1 < \lambda_2$, т.е.

$$\zeta = \zeta_1 - \zeta_2 = \int_0^{T_6} \xi(t)[S_2(t) - S_1(t)]dt > \frac{1}{2}(E_1 + E_2 - 2\sqrt{E_1 E_2} r). \quad (3.11)$$

Случайные величины того типа, к которому принадлежит ζ (интеграл от реализации БГШ с некоторым весом), рассмотрены в 1.2. Было показано, что эта величина является гауссовской, имеет нулевое математическое ожидание (т.е. является центрированной) и дисперсию σ^2 , определяемую соотношением (1.54). В данном случае

$$\sigma^2 = \frac{N_0}{2} \int_0^{T_6} [S_1(t) - S_2(t)]^2 dt = \frac{N_0}{2} (E_1 + E_2 - 2\sqrt{E_1 E_2} r). \quad (3.12)$$

Для нормированной случайной величины $\zeta_n = \zeta/\sigma$ вместо (3.11) получим

$$\zeta_n > \frac{1}{\sqrt{2N_0}} \sqrt{E_1 + E_2 - 2\sqrt{E_1 E_2} r}. \quad (3.13)$$

Наконец, поскольку ζ_n нормальная центрированная случайная величина с единичной дисперсией, то вероятность выполнения неравенства (3.11), т.е. вероятность ошибки при идеальном приеме, составляет

$$P_{\text{ош}} = \Phi \left[\frac{1}{\sqrt{2N_0}} \sqrt{E_1 + E_2 - 2\sqrt{E_1 E_2} r} \right]. \quad (3.14)$$

Последнее выражение полностью определяет потенциальную помехоустойчивость когерентного приема сигналов при двоичной манипуляции. В соответствии с ним помехоустойчивость зависит лишь от энергий ЭС и их взаимной корреляции. Важно, что она не зависит от конкретной формы сигналов и, в частности, от занимаемых ими полос частот. Правда, при двоичной манипуляции полосе уже, чем $1/T_6$ сигналы занимать не могут (это ширина основного спектра отрезка синусоиды длительности T_6). Зато более широкие полосы вполне возможны и в конкретных случаях применяются на практике. Выражение (3.14) показывает, что соответствующее расширение полосы частот, занятых ЭС, само по себе не приводит к какому-либо изменению помехоустойчивости (ни к увеличению, ни к снижению).

В некоторых случаях удобно пользоваться выражением (3.14), представленным в несколько иной форме. Введем понятие расстояния d между двумя сигналами $s_1(t)$, $s_2(t)$, заданными на отрезке времени от a до b , определив его соотношением

$$d = \sqrt{\int_a^b [s_1(t) - s_2(t)]^2 dt} = \sqrt{E_1 + E_2 - 2\sqrt{E_1 E_2} r}. \quad (3.15)$$

Тогда вместо (3.14) получим

$$P_{\text{ош}} = \Phi\left(\frac{d}{\sqrt{2N_0}}\right). \quad (3.16)$$

Остановимся на некоторых частных случаях. Пусть вначале $r = 1$, т.е. оба ЭС совпадают по форме и имеет место амплитудная манипуляция. Тогда

$$P_{\text{ош}} = \Phi\left(\frac{|\sqrt{E_1} - \sqrt{E_2}|}{\sqrt{2N_0}}\right). \quad (3.17)$$

Из выражения (3.17) следует, что при фиксированном значении максимальной мощности ЭС (энергия на бит E_m) наименьшая вероятность ошибки достигается, когда один из ЭС имеет эту мощность, а мощность второго равна нулю. Таким образом, из всех возможных вариантов двоичной АМ наибольшей помехоустойчивостью обладает АМ с пассивной паузой. При этом

$$P_{\text{ош}} = \Phi\left(\sqrt{\frac{E_m}{2N_0}}\right) = \Phi\left(\sqrt{\frac{E}{N_0}}\right) = \Phi(h_6). \quad (3.17a)$$

В (3.17a) E — средняя энергия ЭС; h_6^2 — нормированное отношение сигнал/шум. Последнее определяется как отношение на входе приемника средней энергии сигнала, приходящейся на бит принимаемого сообщения, к спектральной плотности шума:

$$h_6^2 = \frac{E_6}{N_9} = \frac{P_c T_6}{N_0} = \frac{P_c}{\Delta f N_0}, \quad (3.18)$$

где P_c — средняя мощность сигнала на входе приемника; $\Delta f = 1/T_6$ — информационная полоса сигнала. Следует иметь в виду, что информационная полоса — величина условная, определяемая исключительно скоростью передачи информации, не связанная с истинной полосой, занимаемой ЭС. Напомним, что величина h_6 и понятие информационной полосы уже вводились в гл. 2 (выражения (2.28), (2.29)). Нормированное отношение сигнал/шум будет использоваться и в последующих главах, где оно выступает как показатель одной из главных характеристик систем связи — их энергетической эффективности. Имеет смысл более детально обсудить это понятие. Прежде всего, следует обратить внимание на нюансы, относящиеся к приведенным определениям. В гл. 2 термин «энергетическая эффективность» опирался на понятие о бите как единице измерения количества информации, причем предполагалось, что оно измеряется так, как определено теорией, с учетом статистики символов. В выражении (3.18) фактически под битом понимался двоичный символ сигнала, поступающего на вход модулятора, независимо от статистики этих символов. В последующем материале при определении нормированного отношения сигнал/шум бит всегда будет рассматриваться как единица объема передаваемого сообщения (см. гл. 2),

а не количества информации в нем. Возникает следующая проблема: в отличие от количества информации, объем сообщения может меняться при тех или иных его преобразованиях в трактах передачи и приема системы связи. Тогда понятие объема транслируемого по радиолинии сообщения перестает быть однозначным, а вместе с ним исчезает и однозначность определения нормированного отношения сигнал/шум. Ниже во всех разделах книги принято при вычислениях h_6 опираться на объем сообщения на входе рассматриваемых трактов, в том числе и в ситуациях, когда они включают и устройства преобразования сообщения, изменяющие его объем. В последнем случае, если в процессе анализа приходится нормировать отношение сигнал/шум относительно объема преобразованного сообщения, соответствующая величина h_6 снабжается дополнительным индексом, указывающим точку тракта, относительно которой проводится нормирование. Отметим еще, что в элементах трактов радиолиний, анализируемых в настоящей главе, никакие преобразования сообщений, изменяющие их объем, не рассматриваются, так что определение (3.18) является вполне однозначным. Во всех последующих главах, где изучаются вопросы помехоустойчивого кодирования сообщений, являющегося преобразованием, изменяющим их объем, необходимо учитывать сделанное замечание.

Для АМ с пассивной паузой $h_6^2 = E/N_0$, так как ЭС переносит ровно 1 бит, так что его энергия и есть E_6 .

Важный частный случай двоичной передачи образует вариант с нормированными ЭС: $E_1 = E_2 = E$. При этом в соответствии с (3.14) имеем

$$P_{\text{ош}} = \Phi(h_6 \sqrt{1-r}). \quad (3.19)$$

Из последнего выражения следует, что при двоичных нормированных сигналах максимум помехоустойчивости достигается при $r = -1$, т.е. при противоположных сигналах:

$$P_{\text{ош}} = \Phi(\sqrt{2}h_6). \quad (3.19a)$$

Отметим, что это соотношение имеет место независимо от конкретной формы ЭС. В частности, оно относится и к случаю, когда сигналы представляют собой противоположные по фазе отрезки синусоиды, т.е. при ФМ2.

Сопоставляя (3.17a) и (3.19a), убеждаемся, что при противоположных ЭС та же вероятность ошибки, что и для АМ с пассивной паузой достигается при вдвое меньшей средней мощности сигнала. Таким образом, энергетический выигрыш от такого перехода составляет (по средней мощности) 3 дБ.

Наконец, рассмотрим случай, когда ЭС ортогональны ($r = 0$) и нормированы (ортономированные сигналы). Тогда из (3.19) получим

$$P_{\text{ош}} = \Phi(h_6). \quad (3.19b)$$

В частности, этому соотношению удовлетворяет двоичная частотная манипуляция (ЧМ2) при достаточно большом (по сравнению с информационной полосой сигнала) разномес частот, когда ЭС квазиортогональны, т.е. когда $|r| \ll 1$. Сопоставление (3.17a) и (3.19b) позволяет установить, что помехоустойчивость ЧМ2 такая же, как и у АМ с пассивной паузой. Правда, так обстоит дело при сравнении по средней мощности сигнала. Если же интересоваться его пиковой мощностью, то для ЧМ она равна средней, а для АМ в два раза ее превосходит. Поэтому, если сравнивать пиковые мощности ЭС, то при АМ для обеспечения той же вероятности ошибки требуется сигнал в два раза большей мощности, чем при ЧМ.

На рис. 3.3 представлены зависимости, определяющие потенциальную помехоустойчивость простейших двоичных методов манипуляции.

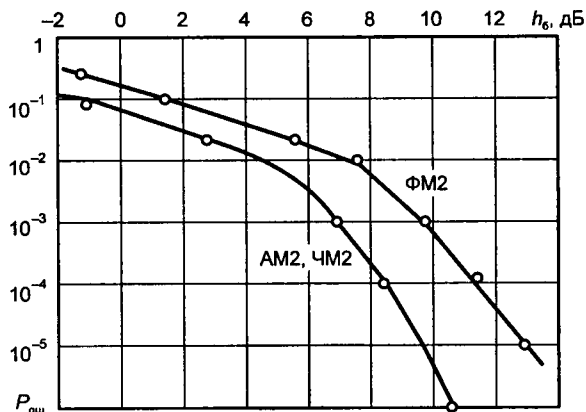


Рис. 3.3. Зависимости, определяющие потенциальную помехоустойчивость базовых методов манипуляции

3.2.3. Потенциальная помехоустойчивость при недвоичной передаче

При m -ичной передаче чаще всего используют $m = 2^k$, где k — целое число. Такой выбор числа сигналов облегчает трансляцию двоичных сообщений, которые, как правило, и встречаются на практике. При этом следующие один за другим двоичные символы сообщения набираются в блоки по k символов и каждому такому блоку ставится в соответствие свой ЭС. Поскольку длительность блока в k раз больше времени передачи бита, то и длительность ЭС оказывается равной $T_3 = kT_6$.

Способ, которым устанавливается соответствие между видом блока символов сообщения и номером ЭС, называется *манипуляционным кодом*. Выбор манипуляционного кода в ряде случаев влияет на помехоустойчивость передачи. Это объясняется следующими обстоятельствами. ЭС могут быть представлены точками в некотором пространстве, причем расстояние между ними может быть введено в соответствии с соотношением (3.15). Как показывает формула (3.16), ошибки демодуляции приводят чаще всего к тому, что вместо действительно переданного ЭС воспроизводится один из ближайших к нему. Это вызывает неверное воспроизведение блока сообщения, что не означает искажения всех двоичных символов, образующих этот блок; часть символов воспроизводится правильно.

Стремясь по возможности расширить эту часть, целесообразно, при установлении соответствия между блоками и ЭС, сигналам с малым расстоянием приписывать мало отличающиеся блоки. Отличие блоков целесообразно оценивать числом позиций в этих блоках, на которых стоят различные символы. Так, блоки 111 и 000 отличаются в трех позициях, так что расстояние между ними примем равным трем. А блоки 110 и 100 отличаются в одной позиции, расстояние между ними примем равным единице. Введенное таким образом расстояние называют *расстоянием Хэмминга*. В отличие от этого расстояние, введенное соотношением (3.15), именуют *среднеквадратичным*. Таким образом, целесообразно так выбирать манипуляционный код, чтобы блокам с меньшим расстоянием Хэмминга соответствовали ЭС с меньшим *среднеквадратичным* расстоянием. Ясно, что при неизменной вероятности ошибочного приема ЭС такой код даст некоторое снижение вероятности ошибки воспроизведения двоичных символов сообщения. К сожалению, такое кодирование может быть достаточно эффективным только, если набор расстояний между используемыми ЭС

подобен набору расстояний Хэмминга между блоками сообщения. А это имеет место, как мы увидим ниже, далеко не всегда. Тем не менее в ряде случаев такое рациональное манипуляционное кодирование (коды Грея) оказывается полезным.

Начнем с рассмотрения двукратной ($k = 2$) фазовой манипуляции ФМ4, когда используются четыре возможных ЭС:

$$s_j(t) = A \cos\left(\omega_0 t + \frac{\pi j}{2}\right), \quad j = 1, 2, 3, 4. \quad (3.20)$$

Эти сигналы представлены точками на векторной диаграмме (рис. 3.4). Около точек записаны соответствующие им блоки. Легко видеть, что манипуляционный код является кодом Грея. Кроме того, на рис. 3.4 проведены две ортогональные оси x и y , повернутые относительно векторов сигналов на 45° и показаны проекции сигналов на данные оси. Длину сигнального вектора в соответствии с (3.15) следует считать равной $\sqrt{P_c}$, где P_c — мощность сигнала. Тогда длины его проекций на ортогональные оси составляют $\sqrt{P_c/2}$. Таким

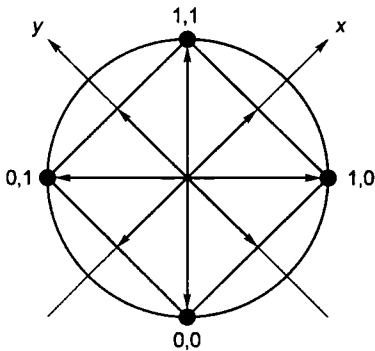


Рис. 3.4. Сигналы и манипуляционный код для ФМ4

образом, передачу каждого ЭС, имеющего мощность P_c , можно рассматривать как одновременную передачу двух квадратурных сигналов, являющихся проекциями ЭС на оси x и y . Мощности каждой такой проекции составляют $P_n = P_c/2$. Отметим теперь важное свойство этих проекций. Пусть на модулятор ФМ4 поступает некоторый блок, состоящий из двух бит, и ему ставится в соответствие некоторый ЭС в соответствии с представленным на рис. 3.4 манипуляционным кодом. Тогда первый бит блока определяет проекцию на ось x , независимо от второго бита, и наоборот, второй бит определяет проекцию на ось y , независимо от первого бита. Соответственно время, в течение которого каждая из проекций остается неизменной, составляет $T_s = 2T_b$. Приведенные данные означают, что передача сигналов ФМ4 с мощностью P_c при манипуляционном коде Грея может интерпретироваться как параллельная передача двух сигналов ФМ2, каждый мощностью $P_2 = P_c/2$. Эти сигналы ФМ2 имеют квадратурные несущие (если для первого канала $\cos\omega_0 t$, то для второго — $\sin\omega_0 t$) и длительности элементарных посылок T_s . Половина информации сообщения (например, нечетные биты: первый, третий и т.д.) передаются через первый канал ФМ2, а все биты на четных местах (второй, четвертый и т.д.) — через второй. Соответственно, битовая скорость передачи D_2 через каждый канал ФМ2 составляет половину битовой скорости D_4 для сигнала ФМ4: $D_2 = D_4/2$. Нормированное отношение сигнал/шум для сигнала ФМ4 составляет $h_6 = \sqrt{\frac{P_c}{D_4 N_0}}$, а для каждого из образующих

его каналов ФМ2 $h_{62} = \sqrt{\frac{P_2}{D_2 N_0}} = \sqrt{\frac{2P_c}{2D_4 N_0}}$. Таким образом,

$$h_{62} = h_6. \quad (3.21)$$

Представление канала с ФМ4 в виде двух параллельных каналов с ФМ2 позволяет попытаться выразить помехоустойчивость ФМ4 через помехоустойчивость ФМ2. Для этого

рассмотрим два, работающих параллельно, идеальных приемника ФМ2, каждый из которых рассчитан на прием одного из квадратурных сигналов совместно образующих сигнал ФМ4. Блок-схема такого устройства представлена на рис. 3.5.

Каждая из двух ветвей предназначена для своего канала ФМ2. Предполагается, что начало отсчета времени в демодуляторе выбрано

так, что $\sin \omega_0 t$ и $\cos \omega_0 t$ совпадают с несущими квадратурных каналов. Интервал интегрирования интеграторов схемы совпадает с тактовым интервалом передачи элементарных сигналов каналов ФМ2 длиной $T_s = 2T_6$. После интеграторов установлены пороговые устройства (ПУ). Они формируют на выходе сигнал «единица», если их входной сигнал положителен и «нуль» — в противном случае. Таким образом, ПУ выполняют роль решающих схем идеальных приемников ФМ2. Нижняя ветвь демодулятора дополнена еще линией задержки (ЛЗ) на время T_6 . Дело в том, что смена выходных бит обоих ПУ происходит одновременно в момент окончания очередного ЭС, так что каждые $2T_6$ секунд появляется сразу два новых бита сообщения. Для того, чтобы выдавать их на выход демодулятора равномерно (по одному каждые T_6 секунд) и необходима ЛЗ. При этом выходной ключ схемы в течение времени T_6 снимает сигнал с верхнего контакта, после чего в течение того же времени с нижнего.

Докажем, что представленный демодулятор действительно является идеальным приемником ФМ4. Ясно, что поскольку квадратурные каналы транслируют различные, не связанные между собой сообщения, то идеальный приемник должен содержать два параллельных тракта, каждый из которых является идеальным приемником для своей квадратуры. Идеальность приемников ФМ2 должна быть доказана при наличии не только БГШ, но и при воздействии сигнала другого квадратурного канала, который выступает уже для данного приемника в качестве помехи. Легко видеть, что на работу каждой из ветвей схемы, представленной на рис. 3.5, сигналы другой ветви вообще не влияют, поскольку в силу ортогональности сигналов обеих квадратур не создают выходного эффекта на выходе интегратора. Таким образом, рис. 3.5 действительно отражает структуру идеального приемника ФМ4. Строго говоря, это справедливо только при условии использования манипуляционного кода Грея.

Теперь не представляет труда определить потенциальную помехоустойчивость ФМ4. Действительно, вероятность ошибки для любой ветви схемы на рис. 3.5, совпадающая с вероятностью ошибки для приемника в целом, в соответствии с (3.19а) составляет

$$P_{\text{ош}} = \Phi(\sqrt{2}h_{62}) = \Phi(\sqrt{2}h_6). \quad (3.22)$$

Здесь использовано соотношение (3.21).

Отметим что полученные данные означают, что потенциальная помехоустойчивость ФМ4 полностью совпадает с потенциальной помехоустойчивостью ФМ2.

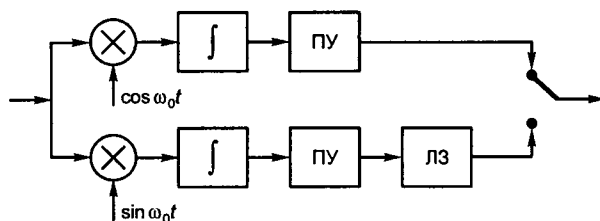


Рис. 3.5. Идеальный когерентный приемник ФМ4: ПУ — пороговое устройство; ЛЗ — линия задержки

3.2.4. Потенциальная помехоустойчивость при ортогональных ансамблях сигналов

Пусть ЭС, которые могут быть преданы на данном такте, образуют ансамбль функций попарно ортогональных на данном интервале. Примем, что все сигналы ансамбля имеют одну и ту же энергию E . Обозначим значность передачи (число сигналов в ансамбле) через m .

Как и выше, предположим, что в действительности был передан первый из этих сигналов, так что $z(t) = S_1(t) + \xi(t)$. Далее можно рассуждать так же как и при выводе соотношений (3.9) и (3.10) с той лишь разницей, что, поскольку энергии всех ЭС одинаковы, можно принять $\lambda_j = \rho_j$. В результате получим

$$\begin{aligned} \lambda_1 &= E + \zeta_1; \\ \lambda_j &= \zeta_j \quad \text{при } j = 2, 3, \dots, m. \end{aligned} \quad (3.23)$$

Все случайные величины ζ_j являются нормальными центрированными случайными величинами с одинаковыми дисперсиями: $\sigma_j^2 = \sigma^2 = EN_0$ (см.(3.12)). В силу ортогональности ЭС эти случайные величины взаимонезависимы.

Правильный прием идеальный когерентный приемник будет осуществлять в тех и только тех случаях, когда

$$\lambda_1 > \lambda_j \quad \text{для всех } j \text{ от } 2 \text{ до } m. \quad (3.24)$$

Вероятность выполнения этого неравенства, т.е. вероятность правильного приема может быть представлена в виде

$$\begin{aligned} P_{\text{пр}} &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} dy \left[\exp - \frac{(y-E)^2}{2\sigma^2} \right] \left\{ \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^y \left(\exp - \frac{x^2}{2\sigma^2} \right) dx \right\}^{m-1} = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left[\exp - 0,5 \left(Y - \frac{E}{\sigma} \right)^2 \right] [1 - \Phi(Y)]^{m-1} dY, \end{aligned} \quad (3.25)$$

где $Y = y/\sigma$.

Пусть теперь $k = \log_2 m$, так что каждый ЭС переносит k бит. Тогда энергия сигнала, расходуемая на передачу одного бита, составляет $E_6 = E/k$, так что вероятность ошибки при приеме ЭС составляет

$$P_{\text{ош}} = 1 - P_{\text{пр}} = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} [\exp - 0,5(Y - \sqrt{k}h_6)^2] [1 - \Phi(Y)]^{m-1} dY. \quad (3.26)$$

Этот интеграл в общем случае не может быть выражен через элементарные функции. Подробный анализ результатов, следующих из этого соотношения, приведен в гл. 4 в материалах, посвященных ортогональному кодированию. Здесь отметим лишь, что соотношение (3.26) при $m \rightarrow \infty$ переходит в (2.32), так что ансамбль ортогональных сигналов представляет собой один из немногочисленных примеров, когда известен конкретный путь реализации пределов, устанавливаемых теорией информации.

3.3. Некогерентный прием

3.3.1. Сигналы со случайной начальной фазой и структура идеального некогерентного приемника

При когерентной демодуляции все параметры ЭС предполагаются априори точно известными. Между тем выше при описании постановки задачи идеального приема отмечалось, что в ряде случаев это требование может не выполняться. Именно, ЭС могут иметь вид $s(\mu, \chi, t)$, причем относительно вспомогательных параметров μ могут быть известны только их априорные распределения вероятностей $p_\mu(x)$. Рассмотрим вопрос об идеальном приеме

в данном случае. Итак, имеется m групп возможных сигналов (номера групп $j = 1, 2, \dots, m$), причем внутри каждой группы сигналы отличаются значениями вспомогательного параметра. Каким в данном случае должен быть идеальный приемник? Для того, чтобы получить наглядный ответ на вопрос, удобно принять, что число возможных значений вспомогательного параметра в каждой группе конечно. Пусть оно составляет n_j , так что общее число возможных ЭС составляет $N = \sum_{j=1}^m n_j$. Пусть априорные вероятности передачи любого из этих

ЭС известны. Тогда, используя подход, развитый в предыдущем разделе, можно определить апостериорную вероятность каждого ЭС (при принятом с выхода канала сигнале $z(t)$). Просуммировав апостериорные вероятности сигналов, входящих в каждую из групп, найдем апостериорные вероятности групп, т.е. апостериорные вероятности того, что в действительности передавался тот или иной символ сообщения. Символ с максимальной такой вероятностью и должен стать выходным символом идеального демодулятора.

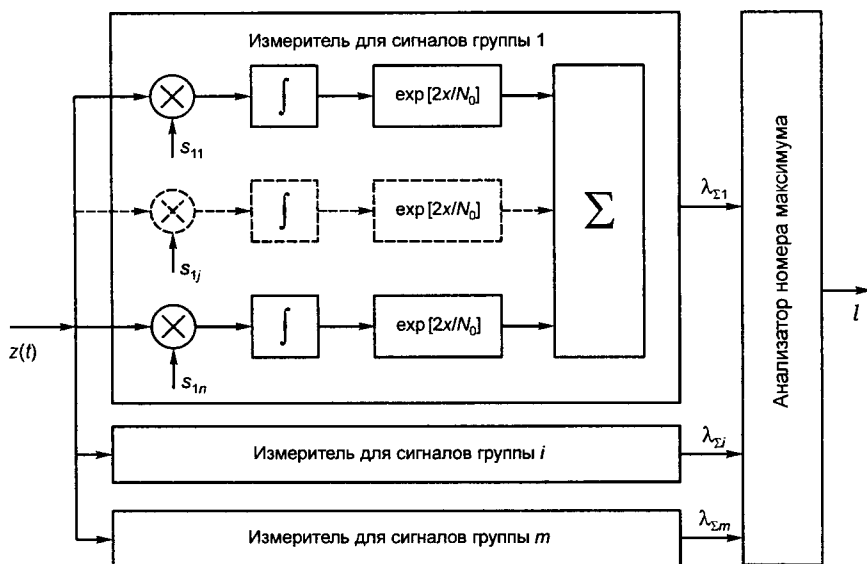


Рис. 3.6. Идеальный приемник для сигналов со случайными параметрами

Примем для простоты, что число сигналов в каждой группе одно и то же ($n_j = n$), априорные вероятности передачи каждого из сигналов равны и их энергии одинаковы. Тогда, как отмечалось выше, апостериорные вероятности могут быть заменены коэффициентами правдоподобия и блок-схема идеального приемника приобретает вид, представленный на рис. 3.6. Под $s_{j\mu}(t)$ понимается μ -й сигнал из j -й группы ($j = 1, 2, \dots, m$, а $\mu = 1, 2, \dots, n$). Представленные на рисунке блоки измерителей определяют коэффициенты правдоподобия $\lambda_{\Sigma j}$ каждой группы сигналов путем суммирования коэффициентов правдоподобия отдельных ЭС, входящих в эту группу. При вычислении последних как и прежде воспользуемся выражением (3.2). Однако здесь уже нельзя заменить истинный вид функции правдоподобия другой, монотонно зависящей от нее функцией, поскольку это может привести к изменению номера той группы, для которой суммарный коэффициент правдоподобия оказывается наибольшим. Что же касается постоянных множителей, то они как и прежде могут быть опущены.

ны. Именно поэтому в схеме на рис. 3.6 появляются новые дополнительные нелинейные элементы с экспоненциальной зависимостью выходного напряжения от входного. Действительно, в соответствии с (3.2) коэффициент правдоподобия для ЭС $s_{j\mu}(t)$ может вычисляться по формуле

$$\lambda_{j\mu} = \exp \left[\frac{2}{N_0} \int_0^{T_s} z(t) s_{j\mu}(t) dt \right], \quad (3.27)$$

а коэффициент правдоподобия каждой группы ЭС —

$$\lambda_{sj} = \sum_{\mu=1}^{n_j} \lambda_{j\mu}. \quad (3.28)$$

По мере роста числа значений вспомогательного параметра μ идеальный приемник становится все более громоздким и в общем случае нереализуемым.

Это не исключает, однако, того, что в отдельных частных случаях необходимые вычисления оказываются не только реализуемыми, но и достаточно компактными. Важнейшим из них является случай, когда начальная фаза радиосигнала μ заранее неизвестна на приемной стороне и распределена равномерно на интервале $(0, 2\pi)$. Идеальный приемник, работающий исходя из такого предположения, носит название *некогерентного*; алгоритм его работы именуют *некогерентным приемом*. Приступим к определению данного алгоритма.

Итак, пусть ЭС имеют вид $S_j(t) = A_j(t) \cos[\omega_0 t + \varphi_j(t) + \mu]$, где ω_0 — несущая частота, а $A_j(t)$ и $\varphi_j(t)$ — амплитуда и фаза ЭС j -й группы. Используя выражение для косинуса суммы углов, находим

$$S_j(t) = S_{jC}(t) \cos \mu + S_{jS}(t) \sin \mu, \quad (3.29)$$

где $S_{jC}(t) = A_j(t) \cos[\omega_0 t + \varphi_j(t)]$ и $S_{jS}(t) = -A_j(t) \sin[\omega_0 t + \varphi_j(t)]$. Указанную пару функций будем называть *квадратурами* j -й группы сигналов.

Тогда в соответствии с (3.27)

$$\lambda_{j\mu} = \exp \left[\frac{2}{N_0} (\lambda_{jC} \cos \mu + \lambda_{jS} \sin \mu) \right], \quad (3.30)$$

где $\lambda_{jC} = \int_0^{T_s} z(t) S_{jC}(t) dt$ и $\lambda_{jS} = \int_0^{T_s} z(t) S_{jS}(t) dt$.

Суммарный коэффициент правдоподобия для j -й группы может быть найден путем усреднения (3.30) по μ :

$$\lambda_{sj} = \frac{1}{2\pi} \int_0^{2\pi} \exp \left[\frac{2}{N_0} (\lambda_{jC} \cos \mu + \lambda_{jS} \sin \mu) \right] d\mu = I_0 \left(\sqrt{\lambda_{jC}^2 + \lambda_{jS}^2} \right), \quad (3.31)$$

где $I_0(x)$ — функция Бесселя мнимого аргумента нулевого порядка. Последний интеграл табличный.

Аргумент функции Бесселя в (3.31) неотрицателен. При этом сама функция монотонно возрастает с ростом аргумента. Поэтому, если заменить ее просто на значение аргумента или его квадрата, что для наших целей удобнее, результат работы схемы на рис. 3.6 не изменится. С учетом этого обстоятельства применительно к нашему случаю каждый измеритель, входящий в состав схемы на рис. 3.6, примет вид, представленный на рис. 3.7.

Этим в сочетании с блок-схемой рис. 3.6 и определится структура идеального некогерентного приема.

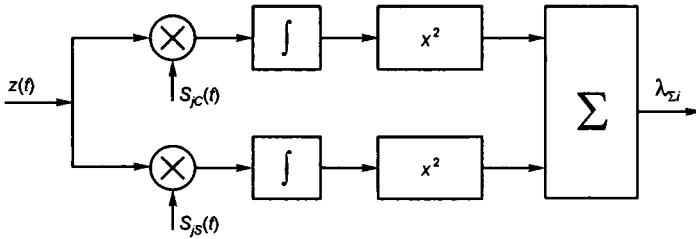


Рис. 3.7. Измеритель коэффициента правдоподобия для сигнала со случайной фазой

3.3.2. Помехоустойчивость некогерентного приема

Ограничимся рассмотрением случая, когда все ЭС имеют равные энергии E и различные группы этих сигналов ортогональны. Последнее требование нуждается в пояснении. Имеется в виду, что любые два сигнала, взятые из различных групп, ортогональны. Иногда говорят, что сигналы ортогональны в усиленном смысле, т.е. при любых начальных фазах. Именно, в соответствии с (3.29) для любых $k \neq l$ и любых значений μ_1 и μ_2 должно выполняться условие

$$\int_0^T [S_{kC}(t) \cos \mu_1 + S_{kS}(t) \sin \mu_1] [S_{lC}(t) \cos \mu_2 + S_{lS}(t) \sin \mu_2] dt = 0. \quad (3.32)$$

Для этого необходимо и достаточно, чтобы любые квадратуры различных групп сигналов были бы попарно ортогональны.

Приведем примеры ортогональных в усиленном смысле сигналов. Пусть рассматривается множество групп сигналов, у которых $\varphi_j(t) = 0$, а $A_j(t)$ попарно ортогональны. Тогда, как следует из (3.32), сигналы ортогональны в усиленном смысле. В качестве ортогонального семейства функций $A_j(t)$ может быть выбрано, например, множество функций Уолша (см. гл. 4) с числом символов, равным значности m . В частном случае двоичной передачи это соответствует использованию двух представленных на рис. 3.8 законов изменения амплитуды. Легко понять, что таким сигналам соответствует двоичная относительная фазовая манипуляция (см. гл. 1). При этом $T_3 = 2T_6$ и ЭС представляет собой пару посылок: предыдущую и очередную; $A_2(t)$ соответствует изменению знака передаваемого сигнала на очередном интервале по сравнению с предыдущим, а $A_1(t)$ — отсутствию такого изменения.

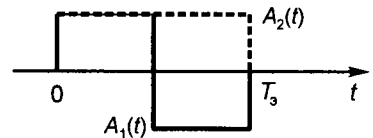


Рис. 3.8. Ортогональные законы изменения амплитуды для двоичной передачи

Другим примером может служить частотная манипуляция. Амплитуда остается постоянной ($A_j(t) = A_j$; $j = 1, 2, \dots, m$), а фаза изменяется по закону $\varphi_j(t) = (j-1)\Omega t$, где Ω — шаг по частоте. Если $\Omega T_3 \gg 2\pi$, то ансамбль сигналов можно считать ортогональным в усиленном смысле.

Переходя к анализу помехоустойчивости, можно без ограничения общности принять, что в действительности был передан некоторый сигнал из первой группы:

$$z(t) = s_{1\mu}(t) + \xi(t), \quad (3.33)$$

где $\xi(t)$ — некоторая реализация БГШ. Примем также, что выходы измерителей всех групп определяются соотношениями $V_j = \sqrt{\lambda_{jC}^2 + \lambda_{jS}^2}$. Поскольку эта величина есть монотонная

функция от введенного ранее суммарного коэффициента правдоподобия группы сигналов, такое определение не изменяет результат идеального приема. В данном случае

$$\lambda_{jC} = \int_0^{T_j} z(t) S_{jC}(t) dt = \int_0^{T_j} s_{1\mu}(t) S_{jC}(t) dt + \int_0^{T_j} \xi(t) S_{jC}(t) dt; \quad (3.34a)$$

$$\lambda_{jS} = \int_0^{T_j} z(t) S_{jS}(t) dt = \int_0^{T_j} s_{1\mu}(t) S_{jS}(t) dt + \int_0^{T_j} \xi(t) S_{jS}(t) dt. \quad (3.34б)$$

Вторые интегралы в двух последних соотношениях (обозначим их через ζ_{jC} и ζ_{jS} соответственно) представляют собой нормальные случайные величины с нулевым математическим ожиданием (центрированные) и дисперсией (см. соотношение (1.54)) $\sigma^2 = \frac{1}{2} N_0 E$. Отметим, что ортогональность ансамбля сигналов позволяет утверждать, что эти случайные величины (все $2m$) взаимно независимы. Первые интегралы в этих выражениях равны нулю для всех $j \neq 1$, что следует из ортогональности сигналов в усиленном смысле. Что же касается $j = 1$, то здесь первые интегралы составляют соответственно $E \cos \mu$ и $E \sin \mu$.

Изложенное позволяет утверждать, что величины V_j представляют собой длины двумерных векторов с нормальными случайными координатами (3.34a) и (3.34б). Известно, что, если эти координаты центрированы ($j \neq 1$), то указанная длина распределена по закону Релея

$$p_j(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right). \quad (3.35)$$

Если не центрированы ($j = 1$) — то по закону Райса

$$p_1(x) = \frac{x}{\sigma^2} I_0\left(\frac{Ex}{\sigma^2}\right) \exp\left(-\frac{x^2 + E^2}{2\sigma^2}\right). \quad (3.36)$$

Правильный прием будет осуществляться тогда и только тогда, когда выходные напряжения всех измерителей окажутся по уровню меньше напряжения на выходе измерителя для первой группы сигналов. Иначе говоря, если уровень на выходе первого измерителя оказался равным некоторой величине x , то для правильного приема необходимо и достаточно, чтобы выходы остальных обнаружителей были бы меньше x . Вероятность этого события (правильного приема ЭС) может быть с использованием (3.35) и (3.36) записана в виде

$$\begin{aligned} P_{\text{nn}} &= \int_0^{\infty} \frac{y}{\sigma^2} I_0\left(\frac{Ey}{\sigma^2}\right) \exp\left(-\frac{y^2 + E^2}{2\sigma^2}\right) \left[\int_0^y \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \right]^{m-1} dy = \\ &= \int_0^{\infty} \frac{y}{\sigma^2} I_0\left(\frac{Ey}{\sigma^2}\right) \exp\left(-\frac{y^2 + E^2}{2\sigma^2}\right) \left[1 - \exp\left(-\frac{y^2}{2\sigma^2}\right) \right]^{m-1} dy. \end{aligned} \quad (3.37)$$

Если в этом соотношении, используя формулу бинома Ньютона, раскрыть квадратные скобки, после чего с помощью табличных интегралов произвести интегрирование, то для вероятности ошибочного приема ЭС найдем

$$P_{\text{ошс}} = 1 - P_{\text{nn}} = \sum_{k=1}^{m-1} (-1)^k C_{m-1}^k \frac{1}{k+1} \exp\left(-\frac{k}{k+1} h^2\right), \quad (3.38)$$

где, как и выше, $h^2 = E/N_0$ — отношение энергии ЭС к спектральной плотности шума. В частном случае двоичной передачи из соотношения (3.38) следует:

$$P_{\text{ошб}} = \frac{1}{2} \exp(-h^2 / 2). \quad (3.39)$$

Для ОФМ, как отмечалось выше, ЭС состоит их двух посылок: данной и предыдущей. Таким образом, $E = 2E_6$, так что $h_6^2 = h^2 / 2$ и

$$P_{\text{ошб}} = \frac{1}{2} \exp - h_6^2. \quad (3.40)$$

Интересно сопоставить этот результат с тем, который можно получить при когерентном приеме ОФМ. Точное значение помехоустойчивости в последнем случае получить не удастся. Однако с точностью, достаточной для наших целей, помехоустойчивость когерентного приема ОФМ можно получить с помощью следующего рассуждения. Сначала будем рассматривать такой сигнал, как ФМ2, принимая его когерентно. Тогда вероятность ошибочного приема посылки определится соотношением (3.19а). Чтобы теперь правильно определить символы, передаваемые методом ОФМ, следует сопоставить значения данной демодулированной посылки и предыдущей и установить, изменилось ли это значение. Ошибка в регистрации этого факта может происходить либо, если ошибочно принят данный символ, либо предыдущий. Поэтому вероятность ошибки в символе ОФМ оказывается приблизительно в два раза больше, чем ФМ, т.е. составляет

$$P_{\text{ошб}} = 2\Phi(\sqrt{2}h_6). \quad (3.41)$$

Используя (3.40) и (3.41), можно вычислить значения нормированного отношения сигнал/шум h_6 , которые необходимо обеспечить для того, чтобы найти то или иное требуемое значение $P_{\text{ошб}}$. Отношение полученных при этом величин h_6 позволяет оценить проигрыш в мощности сигнала ν , происходящий при переходе от когерентного приема к некогерентному. Соответствующие данные приведены в табл. 3.2. Для ЧМ2 аналогичный проигрыш получается путем сопоставления (3.39) и (3.19б).

Таблица 3.2. Энергетические потери при переходе от когерентной демодуляции к некогерентной

$P_{\text{ошб}}$	10^{-1}	$3 \cdot 10^{-2}$	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
ν , дБ, для ОФМ2	0,75	0,75	0,69	0,59	0,5	0,43	0,38
ν , дБ, для ЧМ2	2,93	2,01	1,57	1,1	0,9	0,74	0,48

Данные таблицы показывают, что для достаточно малых вероятностей ошибок когерентный прием, хотя и энергетически более выгоден, чем некогерентный, но его преимущества оказываются сравнительно небольшими. При больших вероятностях ошибок они оказываются более значительными, особенно применительно к ЧМ.

Рассмотрение использования множества ортогональных в усиленном смысле сигналов при двоичной передаче и некогерентном приеме позволяет установить, что неограниченный рост числа сигналов (3.38) приводит к тому же предельному соотношениям, которое имеет место при когерентном приеме (более подробно см. гл. 4).

Глава 4

ПОМЕХОУСТОЙЧИВОЕ КОДИРОВАНИЕ В СИСТЕМАХ ЦИФРОВОЙ РАДИОСВЯЗИ

4.1. Общие принципы помехоустойчивого кодирования и декодирования

Применение кодов, исправляющих ошибки, или помехоустойчивое кодирование является эффективным средством повышения достоверности передачи информации при сохранении неизменными скорости передачи и энергетических параметров канала связи или снижения отношения сигнал/шум, требуемого для обеспечения заданной достоверности приема информации. Так, современные, достаточно просто реализуемые помехоустойчивые коды и алгоритмы их декодирования позволяют для обеспечения общепринятых значений допустимой вероятности ошибки порядка 10^{-5} – 10^{-7} снизить требуемое отношение сигнал/шум (E_b/N_0) на 6–8 дБ по сравнению с передачей без кодирования, т.е. получить эквивалентный энергетический выигрыш от кодирования (ЭВК), равный 6–8 дБ. Во многих случаях такое снижение является экономически эффективным. Так, в космических и спутниковых системах снижение требуемого для обеспечения заданной вероятности ошибки отношения E_b/N_0 на 1 дБ по общепринятым оценкам эквивалентно снижению стоимости космического аппарата на 1 млн долл. США. Более того, каждый выигранный децибел в отношении сигнал/шум в системе связи с низким энергетическим потенциалом может коренным образом расширить и ее область применения.

Теоретически эффективность помехоустойчивого кодирования была доказана Шенноном в работе [4.1], опубликованной в 1948 г. Он показал, что, если скорость передачи информации от некоторого источника меньше пропускной способности канала связи, то соответствующим избыточным кодированием можно обеспечить безошибочную передачу информации через канал связи. Это послужило мощным толчком к развитию теории помехоустойчивого кодирования. Были разработаны различные коды, классы эффективных кодов и алгоритмы их декодирования. Однако применение помехоустойчивого кодирования в реальных системах первоначально сильно отставало. Связано это было с отсутствием в то время адекватной элементной базы, что сильно затрудняло массовую реализацию декодеров. И только появление больших интегральных схем (БИС), особенно специальных БИС кодеров и декодеров и цифровых сигнальных процессоров к середине 1980-х гг. обеспечило чрезвычайно широкое применение помехоустойчивого кодирования в радиотехнических системах передачи цифровой информации различного назначения. Так, в большинстве спутниковых систем применение помехоустойчивого кодирования уже требуется системными стандартами.

Теория помехоустойчивого кодирования основана на использовании достаточно сложного математического аппарата, и особенно, высшей алгебры, изучение которого выходит

за рамки программы радиотехнических вузов. Поэтому цель данного раздела состоит в том, чтобы помочь читателю разобраться в правилах построения современных кодов, наиболее эффективных алгоритмах декодирования и основных принципах их практической реализации с минимальным использованием аппарата высшей алгебры. При необходимости более полное изложение теории и техники помехоустойчивого кодирования и декодирования информации можно найти в работах [4.2, 4.3].

В основе помехоустойчивого кодирования лежит введение в передаваемые сообщения определенной избыточности. Смысл влияния избыточности легко понять из обычной практики. Например, в нашей речи или в тексте вероятны ошибки и, если их количество относительно невелико, то это не мешает правильному восприятию смыслового содержания услышанного или прочитанного сообщения. Объясняется это наличием в сообщениях определенной избыточности, которая позволяет «правильно» восстанавливать полученное сообщение, даже в случае ошибок и определенной связностью символов в словах, а слов в предложениях. Так и при кодировании информации объединение символов в блоки и введение в блоки избыточных (не несущих информацию, но в то же время непосредственно связанных с нею) символов лежит в основе исправления ошибок, возникающих при их передаче по каналу связи. (Объединение символов в блоки при кодировании информации не является необходимым действием и приведено как удобный пример для пояснения.)

Как будет видно из дальнейшего, избыточные символы должны вводиться «разумно» так, чтобы получаемые в результате этого блоки отличались друг от друга наибольшим количеством символов.

Также легко понять, что если абсолютное количество избыточных символов в блоках пропорционально длине блока, т.е. относительное количество избыточных символов в блоках различной длины не зависит от длины блока, то чем больше длина блока, тем больше ошибок в нем может быть исправлено, тем выше помехоустойчивость, обеспечиваемая такими кодами. Но следует отметить, что с ростом длины кода растет сложность (или количество вычислительных операций) декодера и, что возможно еще более важно, задержка в передаче информации получателю на время кодирования и декодирования растет, что для ряда систем может оказаться недопустимым.

Большое влияние на выбор кода и алгоритма его декодирования оказывает канал связи. Ниже рассмотрим наиболее распространенные модели каналов связи. Укрупненная структурная схема цифровой системы связи, являющаяся развитием структурной схемы, представленной в гл. 3 на рис. 3.1, приведена на рис. 4.1.

Функции каждого из элементов, входящих в данную схему, достаточно известны. Коротко отметим, что физический канал представляет собой некоторую среду распространения, в которой действуют помехи, искажающие передаваемые по нему сигналы. Наиболее простой и в то же время широко распространенной (см. гл. 3) признается модель канала связи с помехой в виде БГШ. При этом для описания канала связи используется одна из трех моделей:

- двоичный симметричный канал (ДСК) без памяти;
- дискретный канал без памяти;
- дискретный канал с памятью.

В случае ДСК на входе модулятора и выходе демодулятора имеем только двоичные символы (0 и 1), демодулятор принимает жесткие решения (0 и 1), т.е. мера надежности оценки принимаемого символа фактически не используется. Эту меру надежности называют *метрикой принимаемого символа*. Примером дискретного канала связи без памяти служит канал, для которого в общем случае на входе имеем p -ичные символы, а на выходе q -ичные. Дискретный канал с памятью будет рассмотрен в гл. 8. Наиболее характерным примером

такого канала служит канал с замираниями. Замирания сигнала в таком канале обусловлены меняющимся во времени характером многолучевого распространения волн. Это влечет за собой случайные изменения уровня сигнала. Снижения уровня сигнала (замирания) при помехах типа БГШ приводят к появлению ошибок, группирующихся в пачки. Более подробное описание указанных моделей каналов можно найти, например, в [4.2–4.6].

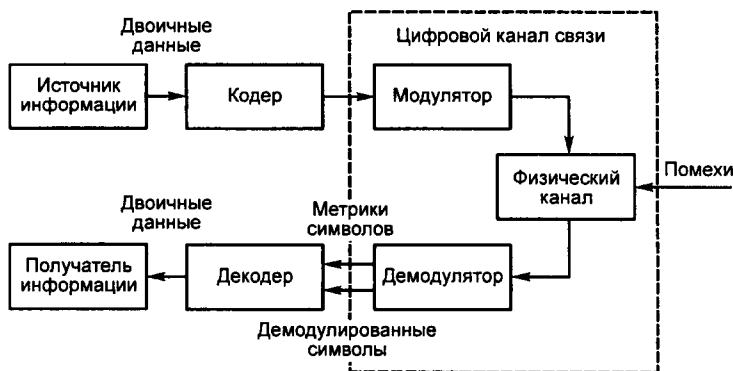


Рис. 4.1. Структурная схема цифровой системы связи

Для каждой модели могут быть предложены наиболее подходящие типы помехоустойчивых кодов.

Кодирование в широком смысле представляет собой преобразование подлежащей передаче последовательности информационных символов $\{a_i\}$ в последовательность кодированных символов $\{b_j\}$. Правило такого преобразования

Таблица 4.1. Пример элементарного избыточного кодирования

a_i	b_j
00	100
01	010
10	001
11	111

может быть задано различным образом. Наиболее наглядным является представление правила преобразования в виде таблицы кодирования. Так, множество всех двухразрядных (кодируемых) комбинаций двоичных последовательностей может быть описано множеством комбинаций трехразрядных (кодированных) двоичных последовательностей (табл. 4.1).

Помехоустойчивое кодирование имеет свои особенности. Здесь, как будет видно из дальнейшего, множество комбинаций кодированных символов должны составлять комбинации, максимально удаленные друг от друга.

Рассмотрим в общем виде основные принципы декодирования.

Пусть на вход физического канала связи поступает сигнал $S(t)$. Из-за действия помех на выходе канала имеем уже другой сигнал $S^*(t)$. Пусть эти сигналы являются элементами некоторого множества \mathcal{B} . Общий принцип декодирования сводится к следующему: на множестве \mathcal{B} вводится метрика, т.е. тем или иным способом определяется расстояние $\mathcal{D}(S_i, S_j)$ между любыми парами сигналов $S_i(t)$ и $S_j(t)$ — элементами данного множества. Декодирование заключается в том, что для принятого сигнала $S^*(t)$ каким-либо способом отыскивается такая комбинация символов $S_j(t)$, для которой $\mathcal{D}(S^*, S_j)$ достигает минимума по всем комбинациям $S_i(t)$. При этом принимается решение о том, что передавался информационный блок, соответствующий кодовой комбинации $S_j(t)$.

Естественно, что осуществляемое таким образом декодирование иногда может приводить к ошибкам. Вероятность ошибки $P_{\text{ошб}}$ при декодировании зависит как от выбора кода, так и от используемой метрики. В теории потенциальной помехоустойчивости показано, что при помехах типа аддитивного белого гауссовского шума при равной априорной вероятности передачи каждого из кодовых слов, наилучшей является определенная в 3.2 среднеквадратичная метрика

$$\mathcal{D}(S, S^*) = \int_0^T [S(t) - S^*(t)]^2 dt. \quad (4.1)$$

Использование именно метрики (4.1) соответствует максимуму помехоустойчивости при прочих равных условиях. Декодер, который декодирует каждую принятую последовательность $S^*(t)$ в ближайшее к ней по минимуму метрики (4.1) кодовое слово $S(t)$, называется *декодером максимального правдоподобия* и в указанных условиях минимизирует вероятность ошибки декодирования принятой последовательности.

Однако нередко желание использовать среднеквадратичную метрику может приводить к достаточно сложной процедуре декодирования с необходимостью вычисления огромного количества операций, поскольку требует вычисления метрики (4.1) для всех возможных кодовых слов $S_i(t)$, где $0 \leq i \leq q^k - 1$, q — основание кода; k — количество информационных символов в кодовой последовательности. Некоторое упрощение процедуры может быть достигнуто путем использования новой метрики, при которой интеграл (4.1) вычисляется не для функций $S(t)$ и $S^*(t)$, а для их цифровых представлений. Наибольшего упрощения можно достичь для двоичных кодов ($q = 2$), когда символы кодовых слов принимаются независимо один от другого и по каждому символу в демодуляторе принимается двоичное решение (0 или 1). Такие двоичные решения на практике называют *жесткими*. Вполне естественно, что использование жестких решений приводит к некоторому снижению помехоустойчивости (снижению энергетического выигрыша от кодирования). Количественная оценка такого снижения будет определена в последующих разделах данной главы.

Принятие демодулятором жестких решений соответствует модели ДСК. Получаемая при таком методе приема метрика равна числу несовпадений символов между принимаемой последовательностью и опорной. Как это было определено в 3.2, такая метрика называется *метрикой Хемминга*, или *хемминговым расстоянием*. Наиболее часто метрика Хемминга применяется при декодировании достаточно большого класса двоичных кодов. Но и использование такой упрощенной метрики при больших k приводит к громоздким схемам или к необходимости выполнения огромного количества вычислений.

Дальнейшего упрощения процедуры декодирования можно достичь, выбирая коды, обладающие специальными алгебраическими свойствами. Как уже отмечалось выше, построение декодера максимального правдоподобия требует вычисления метрик (4.1) для всех возможных кодовых слов, при длинных кодах это требует огромных вычислительных ресурсов. Специальные алгебраические свойства кодов позволяют значительно сократить количество вычислений.

Ниже, наряду с другими наиболее распространенными, будут рассмотрены и коды, обладающие такими алгебраическими свойствами, и алгоритмы их декодирования.

4.2. Классификация помехоустойчивых кодов

Известно большое число помехоустойчивых кодов, отличающихся по структуре, принципам построения, исправляющей способности. Однако все они могут быть принципиально разделены на две большие группы. Первую группу образуют так называемые *блоковые коды*,

вторую — *древовидные*. Определяющее различие между этими кодами состоит в том, что кодер блочного кода — устройство без памяти, отображающее последовательности из k входных символов в последовательности из n выходных. Термин «без памяти» означает, что каждый блок из n выходных символов зависит только от k входных и не зависит от символов других блоков.

Кодер древовидного кода — устройство с памятью. В нем каждый набор выходных символов зависит от текущего набора входных символов и от ν предыдущих входных символов.

Как блочные, так и древовидные коды могут быть систематическими и несистематическими. В кодовых словах систематических кодов передаваемые информационные символы всегда представлены в явном виде.

4.2.1. Блочные коды

Блочный код представляет собой множество последовательностей символов, именуемых *кодowymi словами*. В общем случае элементы кодового слова выбираются из алфавита с q элементами. На практике наиболее широко используются коды, содержащие два элемента (0 и 1), такие коды называются *двоичными* ($q = 2$).

При использовании двоичных блочных кодов непрерывная последовательность информационных бит разбивается на блоки по k бит. Все дальнейшие операции по кодированию и декодированию производятся над каждым блоком отдельно, независимо от других блоков. Каждому такому информационному блоку из k бит ставится в соответствие блок из n символов, где $n > k$. Этот блок из n символов, называемый кодовым словом B_i , передается по цифровому каналу связи, искажается помехами и затем декодируется независимо от других кодовых слов. Величина n называется *длиной кодового слова*. Число различных кодовых слов m в двоичном блочном коде с параметрами (n, k) , очевидно, равно 2^k .

Относительно большой класс блочных кодов составляют линейные коды. Для них по сравнению с общим случаем блочных кодов значительно упрощается операция декодирования. Линейные коды представляют собой подпространство V_q^k линейного пространства V_q^n и обладают следующим важным свойством: сумма (определенная для этого пространства) двух кодовых слов также является кодовым словом. Здесь, как и ранее, q — основание кода; k и n — соответственно количество информационных и кодированных символов в кодовом слове. Для двоичных кодов операция суммирования представляет собой посимвольное сложение двух кодовых слов по модулю 2 (mod 2).

Большим подклассом линейных кодов являются линейные циклические коды (см. 4.4).

К важным параметрам блочного кода относятся скорость кода $R = k/n$, минимальное кодовое расстояние d_{\min} между любой парой кодовых слов данного кода (метрика Хемминга) и вес $w(B)$ кодового слова B , определяемый как число ненулевых символов в кодовом слове B . Можно показать, что для линейных кодов минимальное кодовое расстояние равно расстоянию между нулевым и некоторым другим кодовым словом данного кода, имеющим минимальный вес. Поэтому при построении хороших кодов с максимальным (при прочих равных условиях) d_{\min} следует стремиться к тому, чтобы веса ненулевых кодовых слов были наибольшими. Набор всех весов кода образует распределение (спектр) весов кода, которое обычно используется при нахождении аналитических оценок помехоустойчивости блочных кодов.

Длительность кодового слова равна интервалу времени, в течение которого на входе канала появляется блок из k информационных символов. Декодер принимает решение о переданном кодовом слове после приема всех n символов. (Следует отметить, что при этом декодер должен «знать» начало приема первого символа кодовой последовательности.)

Любой двоичный блоковый код, имеющий минимальное кодовое расстояние $d_{\min} = d$, способен обнаружить любые $f = d - 1$ ошибок или исправить $t = \lfloor (d - 1) / 2 \rfloor$ ошибок. (Здесь $\lfloor x \rfloor$ — целая часть числа x .) Так, код с $d = 3$ способен исправить одну ошибку или обнаружить две. Лучшими среди блоковых кодов будут такие, которые при прочих равных условиях (одинаковых n и k) обладают большим значением d_{\min} , т.е. обнаруживают или исправляют большее количество ошибок.

Среди блоковых кодов можно выделить некоторую группу кодов, которые называются *совершенными* [4.2]. Эти коды позволяют исправить все ошибки кратности t ($t = \lfloor (d - 1) / 2 \rfloor$) и ни одной ошибки большей кратности. По сравнению с другими кодами, имеющими аналогичные значения параметров n и k , кодовые слова совершенных кодов имеют максимальное для выбранных n и k значение d_{\min} , т.е. имеют наибольшую исправляющую способность. Примерами совершенных кодов служат рассматриваемые ниже симплексные коды, а также коды Хемминга и код Голея.

4.2.2. Древоподобные коды

Древоподобные коды это класс кодов, исправляющих ошибки, для которых мгновенные состояния кодера и последовательные их изменения полностью описываются древоподобной диаграммой. Напомним, что при использовании блоковых кодов информационная последовательность разбивается на отдельные блоки, которые кодируются независимо один от другого. Закодированная блоковым кодом последовательность представляет собой последовательность независимых кодовых слов одинаковой длины. В случае древоподобных кодов закодированные символы зависят от ряда предшествующих информационных символов. Поэтому кодер древоподобного кода называют *устройством с памятью*, так что каждый набор n_0 выходных символов такого кодера зависит от текущего набора k_0 входных символов и от v предыдущих входных символов. Память кодера всегда хранит $K = v + k_0$ входных информационных символов. Параметр K называют *длиной кодового ограничения* данного кода. Древоподобные коды характеризуются также скоростью $R = k_0/n_0$ и минимальным свободным расстоянием $d_{св}$. (Понятие минимальное свободное расстояние $d_{св}$ будет определено в 4.5.)

В качестве примера рассмотрим кодер древоподобного кода. Он представляет собой регистр сдвига с одним или несколькими сумматорами. В свою очередь регистр сдвига длины l представляет собой последовательное соединение l ячеек памяти. Состояние ячеек регистра последовательно во времени (в момент прихода каждого нового тактового импульса) меняется так, что состояние ячейки с произвольным номером $i - 1$ переписывается в ячейку с номером i . В кольцевом регистре состояние l -й ячейки (последней в регистре) переписывается в ячейку с номером 1. Связь между ячейками регистра и сумматорами принято описывать, как и в блоковых кодах, порождающими многочленами. Схема одного из наиболее простых кодеров древоподобного двоичного кода с параметрами $R = 1/2$, $K = 3$, заданного двумя многочленами: для верхнего сумматора $G_1(X) = 1 + X^2$ и для нижнего $G_2(X) = 1 + X + X^2$, представлена на рис. 4.2. Для верхнего сумматора в соответствии с многочленом $G_1(X)$ отводы берутся с выходов первой и третьей ячеек регистра, для нижнего — соответственно с первой, второй и третьей. При поступлении на вход данного кодера информационного бита с выхода кодера через коммутатор П сначала с сумматора 1, а затем с сумматора 2 последовательно выходят два закодированных символа.

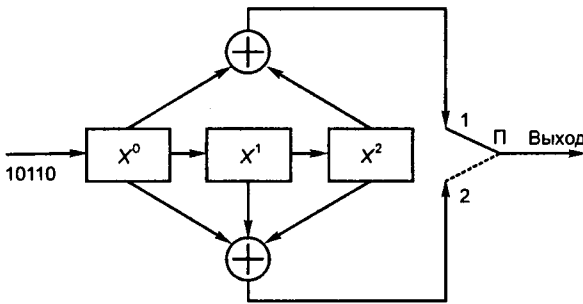


Рис. 4.2. Кодер двоичного древовидного (сверточного) кода с параметрами $R = 1/2$, $v = 2$ и $K = 3$

Древовидная диаграмма для этого кода приведена на рис. 4.3. Вертикальные линии диаграммы называются *ребрами дерева*, горизонтальные — *ветвями*. Каждое ребро дерева соответствует входному символу кодера, каждая ветвь представляет собой последовательность, состоящую из n_0 кодированных (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. 4.3 вершина каждой ветви (узел) обозначена числами от 0 до 3 в соответствии с содержимым двух левых ячеек регистра. Число означает состояние кодера. Отметим, что ребра, выходящие из двух узлов, характеризуемых одинаковым состоянием, тождественны. Проследим, как формируется кодовая (выходная) последовательность символов на примере некоторой произвольной (входной) последовательности информационных бит.

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 кодовой последовательности выделен на рис. 4.3 жирной ломаной линией.

Большой подкласс древовидных кодов составляют сверточные коды и

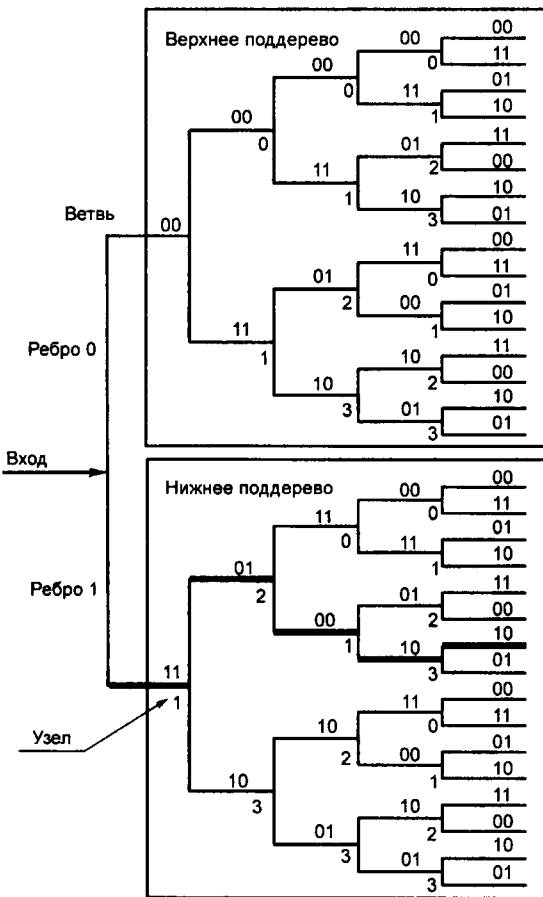


Рис. 4.3. Часть полубесконечного дерева для двоичного древовидного кода с $r = 1/2$, $v = 2$ и $K = 3$

часть турбокодов, которые, в свою очередь, представляют собой подкласс сверточных кодов. Это наиболее широко используемые коды, поскольку для них разработаны высокоэффективные и относительно простые в реализации алгоритмы декодирования.

Выбор кода и алгоритма его декодирования определяется рядом особенностей. К ним, прежде всего, относятся: тип канала связи, достигаемый энергетический выигрыш, допустимая скорость кода (избыточность), задержки на кодирование и декодирование, сложность оборудования (кодера и, особенно, декодера) и др. Более подробно эти вопросы будут рассмотрены в 4.3–4.7.

4.3. Кодирование методом укрупнения алфавита

4.3.1. Оценки помехоустойчивости при когерентном приеме

Данный метод передачи нередко относят к методам m -ичной модуляции (см., например, 3.2). Широко известным примером такой модуляции может служить ранее часто использовавшаяся на практике многопозиционная ЧМ. Однако с полным основанием его относят и к методам кодирования блоковыми кодами (m -ичное кодирование). Это логически очень просто, но в то же время весьма эффективный метод помехоустойчивого кодирования. При $m \rightarrow \infty$ такое кодирование позволяет сколь угодно приблизиться к пределу Шеннона ($E_b/N_0 = \ln 2$) при неограниченно большой полосе канала связи. Ниже это будет продемонстрировано на конкретных примерах.

Операция кодирования состоит в отображении блока из k последовательных независимых двоичных символов в один из $m = 2^k$ подлежащих передаче сигналов, т.е. происходит переход от передачи символов двоичного алфавита к передаче символов (кодовых слов) m -ичного алфавита. Причем символы m -ичного алфавита должны приниматься в целом. Упрощенная функциональная схема приемника сигналов m -ичного алфавита приведена в 3.2. Однако указанная выше высокая эффективность такого кодирования может быть реализована только в случае m ортогональных или симплексных сигналов. (По определению эти сигналы — равноудаленные.) При анализе помехоустойчивости при кодировании укрупнением алфавита следует учитывать два обстоятельства, по-разному влияющие на итоговый показатель помехоустойчивости:

- энергия символа укрупненного алфавита E_m в k раз больше энергии бита, это 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. Это ведет к снижению помехоустойчивости.

Для количественной оценки помехоустойчивости при m -ичном кодировании и приеме по критерию максимального правдоподобия воспользуемся без вывода рядом известных результатов.

Ортогональные сигналы. Вероятность правильного приема на фоне помехи типа БГШ ортогональных кодовых слов m -ичного алфавита определяется выражением (3.25). Каждое из принимаемых при такой передаче кодовых слов декодируется в k информационных бит. Если произошла ошибка в приеме m -ичного символа, то вероятность того, что будут искажены l бит из k переданных, равна

$$C_k^l / \sum_{l=1}^k C_k^l, \quad (4.2)$$

где C_k^l — число сочетаний из k по l . Используя (4.2), определим среднее число искаженных бит, если искажена последовательность из k бит:

$$\sum_{l=1}^k l C_k^l / \sum_{l=1}^k C_k^l = \frac{k 2^{k-1}}{m-1}, \quad (4.3)$$

а условная вероятность искажения данного бита, если искажена последовательность из k бит, внутри которой он передавался, составляет

$$\frac{m}{2(m-1)}. \quad (4.4)$$

Учитывая (4.4), вероятность ошибки в приеме бита при когерентном приеме на фоне помехи типа БГШ с использованием метода помехоустойчивого кодирования m -ичными ортогональными сигналами определяется следующим выражением:

$$P_{\text{ошб}} = \frac{m}{2(m-1)} \left(1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \left[\Phi \left(x + \sqrt{\frac{2E_m}{N_0}} \right) \right]^{m-1} dx \right), \quad (4.5)$$

где E_m — энергия m -ичного кодового слова.

В частном случае, когда $k = 1$, а $m = 2$, из выражения (4.5) следует широко известное выражение

$$P_{\text{ошб}} = \Phi \left(\sqrt{\frac{E_b}{N_0}} \right). \quad (4.6)$$

Вычисления по формуле (4.5) возможны только численными методами. Графики зависимости вероятности ошибки на бит от отношения E_b/N_0 при когерентном приеме m -ичных ортогональных кодовых слов, заимствованные из работы [4.4], приведены на рис. 4.4.

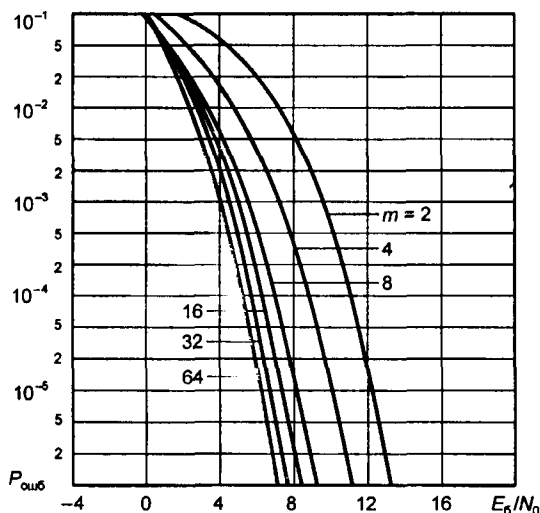


Рис. 4.4. Зависимости вероятности ошибки на бит от отношения E_b/N_0 (когерентный прием)

Из рис. 4.4 видно, что с увеличением размерности алфавита m требуемое для обеспечения заданной вероятности ошибки отношение E_b/N_0 существенно уменьшается. Так, для наиболее часто применяемых значений вероятности ошибки $P_{\text{ошб}} = 10^{-5}$ при $m = 8$ имеем выигрыш от кодирования порядка 4 дБ, а при $m = 64$ выигрыш от кодирования составит 6 дБ. Скорость роста выигрыша от кодирования с дальнейшим увеличением m уменьшается. Однако, несмотря на это, при $m \rightarrow \infty$ требуемое для сколь угодно малой вероятности ошибки значение E_b/N_0 стремится сверху к значению $\ln 2 = -1,59$ дБ. Это минимальное значение $E_b/N_0 = \ln 2$ (см. гл. 2), является пределом Шеннона для канала с бесконечной полосой пропускания и помехами типа аддитивного БГШ.

Простой алгоритм построения ансамблей ортогональных сигналов. Известны различные алгоритмы построения ансамблей (множеств) ортогональных цифровых сигналов. Общим условием ортогональности сигналов $s_i(t)$ и $s_j(t)$ является равенство нулю коэффициента корреляции

$$r = \int_0^T s_i(t) \times s_j(t) dt = 0 \text{ при } i \neq j. \quad (4.7)$$

Рассмотрим достаточно универсальный метод построения ансамбля ортогональных сигналов в виде двоичных кодовых последовательностей. Ансамбль таких последовательностей строится на основе матриц Адамара.

Матрица Адамара первого порядка ($k = 1, m = 2$) имеет следующий вид:

$$\mathbf{H}_1 = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix}. \quad (4.8)$$

Матрица Адамара второго порядка ($k = 2, m = 4$):

$$\mathbf{H}_2 = \begin{vmatrix} \mathbf{H}_1 & \mathbf{H}_1 \\ \mathbf{H}_1 & -\mathbf{H}_1 \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{vmatrix}, \quad (4.9)$$

где все элементы матрицы $-\mathbf{H}_1$ противоположны элементам матрицы \mathbf{H}_1 .

По аналогии с построением матриц \mathbf{H}_1 и \mathbf{H}_2 строится матрица Адамара k -го порядка ($m = 2^k$)

$$\mathbf{H}_k = \begin{vmatrix} \mathbf{H}_{k-1} & \mathbf{H}_{k-1} \\ \mathbf{H}_{k-1} & -\mathbf{H}_{k-1} \end{vmatrix}. \quad (4.10)$$

Строки матриц, построенных во времени по правилам (4.8)–(4.10), представляют собой ансамбль m ортогональных кодовых слов (сигналов). Обычно их называют *функциями Уолша*.

Для любой пары ортогональных кодовых слов из ансамбля (4.10) половина двоичных символов совпадает, половина противоположна. Новая система из m ортогональных сигналов может быть построена перемножением произвольной двоичной последовательности длины m с каждой строкой матрицы порядка $\log_2 m$.

Ансамбль из $m = 2^k$ биортогональных сигналов может быть построен из некоторого множества, содержащего 2^{k-1} ортогональных сигналов, путем добавления к нему 2^{k-1} сигналов противоположных сигналам исходного ансамбля

$$\mathbf{H}_{\text{би}k} = \begin{vmatrix} \mathbf{H}_{k-1} \\ -\mathbf{H}_{k-1} \end{vmatrix}. \quad (4.11)$$

Так для $k = 3$ из (4.11) имеем

$$\mathbf{H}_{\text{би}3} = \begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ -1 & -1 & -1 & -1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 \end{vmatrix}. \quad (4.12)$$

В матрице (4.12) попарно противоположными являются 1-я и 5-я, 2-я и 6-я и т.д. строки.

При $m = 2$ ансамбль биортогональных сигналов представляет собой два противоположных сигнала. В этом случае вероятность ошибки при $M = 2$ определяется известным выражением

$$P_{\text{ошб}} = \Phi \left(\sqrt{\frac{2E_b}{N_0}} \right). \quad (4.13)$$

Вероятность ошибки при когерентном приеме на фоне помехи типа БГШ биортогональных сигналов m -ичного алфавита для $m \geq 8$ практически совпадает со случаем использования m -ичного алфавита ортогональных сигналов.

Симплексные сигналы. Ансамбль из $m = 2^k$ симплексных сигналов (кодовых слов) может быть построен на основе матрицы Адамара порядка k удалением из него первого столбца, в котором все элементы одинаковы. Поэтому для любой пары таких кодовых слов $(m/2) - 1$ двоичных символов, стоящих на одинаковых позициях, совпадают, а $m/2$ не совпадают. Коэффициент корреляции r любой пары кодовых слов такого ансамбля отрицателен:

$$r = \frac{-1}{m-1}. \quad (4.14)$$

Это также подтверждает тот факт, что при $m = 2$ симплексные сигналы оказываются противоположными.

Следовательно, кодирование симплексными последовательностями при одном и том же значении m позволяет обеспечить одинаковую вероятность ошибки, имея значение E_b/N_0 в $(m-1)/m$ меньше, чем требуется для ортогональных последовательностей. При больших значениях m разница становится пренебрежимо малой, и все три типа систем кодовых последовательностей: ортогональных, биортогональных и симплексных обеспечивают практически одинаковую помехоустойчивость и практически одинаковый энергетический выигрыш от кодирования. В то же время для каждой из таких систем требования к полосе пропускания канала и сложность устройств декодирования увеличиваются пропорционально росту m , так что m -ичное кодирование применимо лишь тогда, когда имеется значительный запас полосы относительно полосы, необходимой для передачи некодированной информации. Отметим, что при использовании биортогональных систем полоса и сложность растут пропорционально $m/2$.

4.3.2. Оценки помехоустойчивости при некогерентном приеме

Известно, что в основе некогерентного приема при помехах типа БГШ лежит вычисление распределений огибающих шума и смеси сигнала и шума. Поэтому при m -ичном кодировании оптимальными, т.е. наиболее удаленными, будут ортогональные сигналы.

Вероятность правильного приема при некогерентном приеме на фоне помехи типа БГШ ортогональных (сигналов) кодовых слов m -ичного алфавита определяется выражением (3.37). Используя это выражение и значение условной вероятности искажения любого бита, если искажена последовательность из k бит, внутри которой он передавался (выражение 4.11), по аналогии с выводом (4.5) получим выражение для вероятности ошибки в приеме бита при некогерентном приеме:

$$P_{\text{ошб}} = \frac{m}{2(m-1)} \cdot P_{\text{ошз}}. \quad (4.15)$$

В частном случае при $m = 2$ имеем известное выражение

$$P_{\text{ошб}} = \frac{1}{2} \exp(-E_b/(2N_0)). \quad (4.16)$$

Графики зависимости вероятности ошибки на бит от отношения E_b/N_0 при некогерентном приеме m -ичных ортогональных кодовых слов приведены на рис. 4.5.

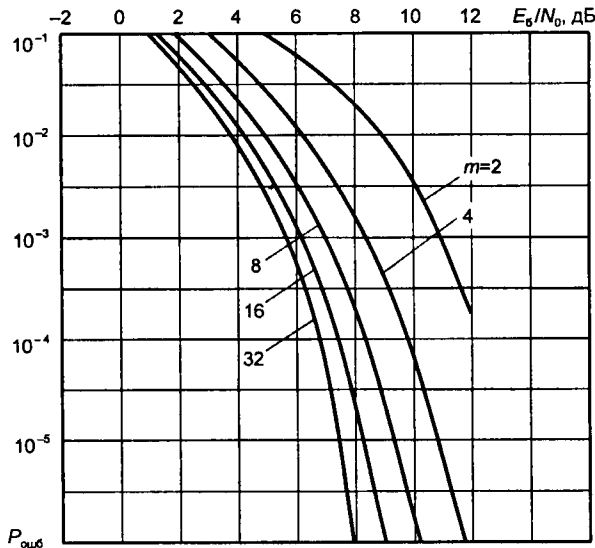


Рис. 4.5. Зависимости вероятности ошибки на бит от отношения E_b/N_0 (некогерентный прием)

Из графиков видно, что с увеличением m , как и при когерентном приеме требуемое для обеспечения заданной вероятности ошибки отношение E_b/N_0 существенно уменьшается и также стремится к пределу Шеннона. Потери E_b/N_0 при некогерентном приеме в целом по сравнению с когерентным с ростом m становятся пренебрежимо малыми.

Для значений вероятности ошибки порядка 10^{-5} и $m = 16, 32$ потери при некогерентном приеме составляют менее 0,2 дБ. При малых значениях E_b/N_0 , соответствующих значениям $P_{\text{ошб}} \leq 10^{-2} - 10^{-3}$, потери при некогерентном приеме возрастают.

4.4. Циклические коды и методы их декодирования

Важным и многочисленным классом часто используемых на практике блочных кодов являются циклические коды. Поиск лучших среди блочных кодов, т.е. таких, которые при прочих равных условиях (одинаковых n и k) обнаруживают или исправляют наибольшее количество ошибок, связан с наложением 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 определенных алгебраических структур, которые имеют своими элементами символы, содержащиеся в кодовых словах.

4.4.1. Определения основных алгебраических структур

Приводимые ниже определения алгебраических структур дают только минимальные предпосылки для изучения блоковых кодов и алгоритмов их декодирования. Определим основные алгебраические структуры: группы, кольца и поля.

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

а) операция может быть применена к двум любым элементам группы, в результате чего получается третий элемент этой же группы;

б) для любых трех элементов группы a , b и c $(a + b) + c = a + (b + c)$, если определена операция сложения и $(ab)c = a(bc)$, если определена операция умножения;

в) в группе существует единственный (нейтральный) единичный элемент. Если определена операция сложения, то единичный элемент обозначается «0» и определяется из уравнения $0 + a = a + 0 = a$. Если определена операция умножения, то единичный элемент обозначается «1» и определяется из уравнения $1a = a1 = a$;

г) каждый элемент группы имеет единственный обратный элемент. Элемент, обратный элементу группы a , обозначается $-a$. Если определена операция сложения, то обратный элемент определяется из уравнения $a + (-a) = (-a) + a = 0$. Если определена операция умножения, то обратный элемент обозначается a^{-1} и определяется из уравнения $a a^{-1} = a^{-1}a = 1$.

Если для элементов группы определена операция сложения и выполняется равенство $a + b = b + a$ или определена операция умножения и выполняется равенство $ab = ba$, то такая группа называется *абелевой*. По операции сложения абелева группа еще называется *аддитивной абелевой*. В качестве наиболее понятного примера укажем, что совокупность всех действительных чисел является группой относительно операции обычного сложения, совокупность всех действительных чисел без нуля является группой относительно операции обычного умножения.

Кольцом R называется множество элементов или символов, для которых определены две операции. Одна из них называется сложением и обозначается как $a + b$, другая умножением и обозначается как ab . Для того, чтобы R было кольцом, должны выполняться следующие аксиомы:

а) множество R является аддитивной абелевой группой относительно операции сложения;

б) для любых двух элементов a и b из множества R определено произведение ab , которое является элементом R ;

в) для любых трех элементов a , b и c из множества R $a(bc) = (ab)c$;

г) для любых трех элементов a , b и c из множества R справедливы равенства $a(b + c) = ab + ac$ и $(b + c)a = ba + ca$.

Кольцо называется коммутативным, если коммутативна операция умножения, т.е. для любых двух его элементов a и b справедливо равенство $ab = ba$.

Поле F называется множество элементов, в котором определены обе операции — сложение и умножение, причем это множество по операции сложения образует аддитивную абелеву группу с единственным элементом, обозначенным «0»; по операции умножения группу образуют все элементы поля, кроме 0, единичный элемент этой группы обозначается «1». Так множество действительных или комплексных чисел являются полями.

Известно, что 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, называются *полями Галуа* и обозначаются $\mathbf{GF}(q)$. Наиболее простую структуру имеют поля Галуа при $i = 1$. Тогда элементам поля удобно поставить в соответствие натуральные числа от 0 до $p - 1$ и операции сложения и умножения

производить над этими числами по mod p . Такие поля Галуа называются *простыми*. При $i > 1$ поля Галуа являются расширениями простого поля. Примером такого поля Галуа $\mathbf{GF}(q)$, где $q = p^l$, является множество всех многочленов степени $l - 1$ или менее, коэффициенты которого являются элементами простого поля Галуа $\mathbf{GF}(p)$.

Именно поля Галуа будут представлять наибольший интерес для дальнейшего описания блоковых кодов.

4.4.2. Циклические коды

Как уже было указано выше, большим подклассом линейных блоковых кодов, наиболее широко используемых на практике, являются циклические коды. Это объясняется значительным упрощением процедур кодирования и декодирования за счет использования алгебраических и циклических свойств этих кодов.

Циклическим кодом называется такой блоковый код, которому вместе с некоторым кодовым словом принадлежат и другие кодовые слова, являющиеся всеми его циклическими сдвигами. При этом следует отметить, что не все кодовые слова циклического кода представляют собой циклические сдвиги друг друга. Пусть $B = [b_{n-1} b_{n-2} \dots b_1 b_0]$ — кодовое слово циклического (n, k) кода, тогда, согласно определению, этому коду принадлежит и любой его циклический сдвиг. Например, последовательность $B^{(1)} = [b_{n-2} \dots b_1 b_0 b_{n-1}]$, являющаяся циклическим сдвигом на один символ кодового слова B , также представляет собой кодовое слово данного циклического кода.

При операциях с циклическими кодами обычно кодовое слово $B = [b_{n-1} b_{n-2} \dots b_1 b_0]$ удобно представлять в виде многочлена $B(X)$ степени $\leq n - 1$, где x — некоторая абстрактная переменная, а коэффициенты при степенях переменной этого многочлена — символы кодового слова:

$$B(X) = b_{n-1} X^{n-1} + b_{n-2} X^{n-2} + \dots + b_1 X + b_0. \quad (4.17)$$

Коэффициенты многочлена $B(X)$ это элемент поля Галуа.

Многочлен, соответствующий кодовому слову $B^{(1)} = [b_{n-2} \dots b_1 b_0 b_{n-1}]$, представляющий собой циклический сдвиг на один символ кодового слова B

$$B^{(1)}(X) = b_{n-2} X^{n-1} + b_{n-3} X^{n-2} + \dots + b_1 X^2 + b_0 X + b_{n-1}, \quad (4.18)$$

можно получить умножением многочлена $B(X)$ на X , т.е. $B^{(1)}(X)$ — остаток от деления произведения $[X B(X)]$ на $(X^n - 1)$. По аналогии кодовое слово $B^{(i)}(X)$ циклического кода с номером i получается как остаток от деления произведения

$$[X^i B(X)] \text{ на } (X^n - 1), \text{ где } i = 0, 1, \dots, n - 1. \quad (4.19)$$

Для двоичного кода каждый из коэффициентов многочленов (4.17)–(4.19) является нулем или единицей.

Определим также многочлен $C(X)$ информационного сообщения, состоящего из k символов:

$$C(X) = c_{k-1} X^{k-1} + c_{k-2} X^{k-2} + \dots + c_1 X + c_0, \quad (4.20)$$

где последовательность $C = [c_{k-1} c_{k-2} \dots c_1 c_0]$ определяет k информационных символов.

Для того чтобы построить двоичный циклический (n, k) код, введем некоторый многочлен $g(X)$ степени $n - k$ с двоичными коэффициентами, который является одним из сомножителей при разложении многочлена $X^n - 1$ на множители:

$$g(X) = X^{n-k} + g_{n-k-1} X^{n-k-1} + \dots + g_1 X + 1. \quad (4.21)$$

Многочлен $g(X)$ называют *порождающим*.

Способ определения порождающих многочленов для различных n и k требует более подробного изучения свойств алгебраических структур, что выходит за рамки данной работы. Для практического использования описанного алгоритма построения двоичных циклических (n, k) кодов и выбора порождающих многочленов можно воспользоваться табл. 4.2, в которой приведены в восьмеричной форме коэффициенты порождающих многочленов для двоичных циклических кодов длиной $n = 7, 15, 31, 63$ и 127 . Более объемный список порождающих многочленов дан в [4.2].

Вполне очевидно, что произведение многочленов $C(X)$ и $g(X)$ это многочлен степени не большей $n - 1$; может представлять собой некоторое кодовое слово циклического (n, k) кода. Поскольку имеется q^k различных информационных сообщений (блоков) размерности k , следовательно, имеется и q^k кодовых слов, которые можно формировать в виде произведения многочленов $C^{(i)}(X)$ и $g(X)$. Обозначим эти кодовые слова как

$$B^{(i)}(X) = C^{(i)}(X)g(X), \quad (4.22)$$

где $i = 0, 1, \dots, q^k$.

Для подтверждения того, что циклические сдвиги (4.22) являются кодовыми словами, рассмотрим какое-либо кодовое слово $B(X)$. Циклический сдвиг кодового слова $B(X)$ можно представить в виде

$$B^{(1)}(X) = XB(X) + b_{n-1}(X^n - 1) \quad (4.23)$$

и, поскольку по определению $X^n - 1$ и $B(X)$ делятся на $g(X)$ без остатка, то и $B^{(1)}(X)$ делится на $g(X)$ без остатка, т.е. $B^{(1)}(X)$ можно представить:

$$B^{(1)}(X) = C^{(1)}(X)g(X). \quad (4.24)$$

Следовательно, циклический сдвиг кодового слова $B(X)$, формируемого по правилу (4.22) с использованием порождающего многочлена $g(X)$, порождает другое кодовое слово того же циклического (n, k) кода.

В качестве примера построим двоичный циклический код с длиной блока $n = 7$. Как следует из табл. 4.2, для выбранного значения $n = 7$ имеется порождающий многочлен 13, на основе которого можно сформировать код (7,4). Обозначению 13 в восьмеричном представлении соответствует двоичное представление 001 011. Соответствующий этой символикe порождающий многочлен примет вид

$$g(X) = X^3 + X + 1. \quad (4.25)$$

Пусть имеем информационное сообщение $C(X)$ в виде произвольной k -разрядной ($k = 4$) двоичной последовательности, например 0 0 1 1. Его, согласно (4.20), можно представить в виде многочлена $C(X) = X + 1$.

Перемножая $C(X)$ и $g(X)$, получим многочлен $X^4 + X^3 + X^2 + 1$, которому будет соответствовать кодовое слово 0 0 1 1 1 0 1.

Как было показано выше, порождающий многочлен $g(X)$ является делителем двучлена $X^n - 1$. Следовательно, двучлен $X^n - 1$ можно представить в виде произведения

$$X^n - 1 = g(X) h(X). \quad (4.26)$$

Многочлен $h(X)$ в (4.26) называют *проверочным*, его степень равна k , поскольку степень порождающего многочлена для кода (n, k) равна $n - k$. Проверочный многочлен $h(X)$, соответствующий порождающему $g(X) = X^3 + X + 1$, имеет следующий вид:

$$h(X) = X^4 + X^2 + X^1 + 1. \quad (4.27)$$

Таблица 4.2. Коэффициенты порождающих многочленов (в восьмеричной форме) некоторых двоичных циклических кодов

n	k	t	$g(X)$
7	4	1	13
15	11	1	23
	7	2	721
	5	3	2467
31	26	1	45
	21	2	3551
	16	3	107657
	11	5	5423325
	6	7	313365047
63	57	1	103
	51	2	12471
	45	3	1701317
	39	4	166623567
	36	5	1033500423
	30	6	157464165547
	24	7	17323260404441
	18	10	1363026512351725
	16	11	6331141367235453
	10	13	472622305527250155
	7	15	5231045543503271737
127	120	1	211
	113	2	41567
	106	3	11554743
	99	4	3447023271
	92	5	624730022327
	85	6	130704476322273
	78	7	26230002166130115
	71	9	6255010713253127753
	64	10	1206534025570773100045
	57	11	335265252505705053517721
	50	13	54446512523314012421501421
	43	14	17721772213651227521220574343
	36	15	3146074666522075044764574721735
	29	21	403114461367670603667530141176155
	22	23	123376070404722522435445626637647043
	15	27	22057042445604554770523013762217604353
	8	31	70472640527510306514762242715677331330217

Примечание. Здесь t — число ошибок в кодовом слове, которые гарантированно исправит данный код.

Прямым умножением легко показать, что произведение $g(X)h(X) = X^n - 1$.

Используя проверочный многочлен $h(X)$ в качестве порождающего, можно построить $(n, n - k)$ циклический код, дуальный исходному коду с порождающим многочленом $g(X)$.

Кодовые слова, генерируемые многочленом $g(X) = X^3 + X^1 + 1$, приведены в табл. 4.3.

Таблица 4.3. Кодовые слова циклического кода (7, 4)

№ п/п	Информационные биты				Кодовые слова						
	X^3	X^2	X^1	X^0	X^6	X^5	X^4	X^3	X^2	X^1	X^0
1	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	1	0	0	0	1	0	1	1
3	0	0	1	0	0	0	1	0	1	1	0
4	0	0	1	1	0	0	1	1	1	0	1
5	0	1	0	0	0	1	0	1	1	0	0
6	0	1	0	1	0	1	0	0	1	1	1
7	0	1	1	0	0	1	1	1	0	1	0
8	0	1	1	1	0	1	1	0	0	0	1
9	1	0	0	0	1	0	1	1	0	0	0
10	1	0	0	1	1	0	1	0	0	1	1
11	1	0	1	0	1	0	0	1	1	1	0
12	1	0	1	1	1	0	0	0	1	0	1
13	1	1	0	0	1	1	1	0	1	0	0
14	1	1	0	1	1	1	1	1	1	1	1
15	1	1	1	0	1	1	0	0	0	1	0
16	1	1	1	1	1	1	0	1	0	0	1

В качестве примера рассмотрим алгоритм формирования систематического p -ичного циклического кода с использованием порождающего многочлена $g(X)$. Пусть имеем многочлен информационного сообщения $C(X)$ степени не более $k - 1$. Умножив его на X^{n-k} , получим

$$X^{n-k}C(X) = c_{k-1}X^{n-1} + c_{k-2}X^{n-2} + \dots + c_1X^{n-k+1} + c_0X^{n-k}. \quad (4.28)$$

Этот многочлен определяет первые k символов кодового слова $B(X)$, соответствующего информационному сообщению $C(X)$. Для того, чтобы получить соответствующее информационному сообщению $C(X)$ кодовое слово, необходимо добавить к многочлену $X^{n-k}C(X)$ многочлен степени меньше, чем $n - k$ и представляющий собой проверочные символы. Для их формирования разделим $X^{n-k}C(X)$ на порождающий многочлен $g(X)$. В результате деления получим

$$\frac{X^{n-k}C(X)}{g(X)} = Q(X) + \frac{r(X)}{g(X)}, \quad (4.29)$$

где $Q(X)$ — частное, а $r(X)$ — остаток от деления, имеющий степень меньше, чем $n - k$. Умножив обе части выражения (4.29) на $g(X)$, получим

$$X^{n-k}C(X) = Q(X)g(X) + r(X). \quad (4.30)$$

Остаток $r(X)$ можно представить полиномом $r_0 + r_1X + r_2X^2 + \dots + r_{n-k-1}X^{n-k-1}$. Из выражений (4.24) и (4.30) следует, что $Q(X)g(X)$ — кодовое слово циклического (n, k) кода. Следовательно, суммируя по $\text{mod } p$ $r(X)$ с $X^{n-k}C(X)$, получим желаемый p -ичный циклический (n, k) код.

Из двоичного циклического (n, k) кода можно построить линейный двоичный расширенный $(n + 1, k)$ и укороченный $(n - l, k - l)$ коды. Отметим, что это положение справедливо не только для циклических, но и для всех линейных двоичных кодов

Расширенный $(n + 1, k)$ код синтезируется путем добавления к каждому кодовому слову (n, k) кода еще одного символа проверки на четность, нуля, если в кодовом слове четное число единиц и единицы, если число единиц нечетное. Это приводит к увеличению минимального расстояния между кодовыми словами расширенного кода на единицу и, как следствие, увеличению исправляющей способности расширенного кода.

Укороченный $(n - l, k - l)$ код строится из систематического линейного двоичного кода путем размещения в начале каждого кодового слова последовательности из l нулей. Эти l нулей далее по каналу связи не передаются, а $n - k$ проверочных символов формируются так же, как и в исходном коде. Укороченный $(n - l, k - l)$ код содержит 2^{k-l} кодовых слов. Минимальное расстояние этих кодовых слов не меньше, чем минимальное расстояние между словами исходного (n, k) кода.

В ряде случаев оказывается более удобным при реализации кодера циклического кода на основе регистра сдвига пользоваться проверочным многочленом $h(x)$. Также проверочный многочлен, как будет видно из дальнейшего, используют при построении декодеров, реализующих описанные ниже мажоритарные алгоритмы декодирования. В этом случае на основе проверочного многочлена $h(x)$ формируются $n - k$ проверочных соотношений вида

$$S_{j-k} = \sum_{i=0}^k h_i b_{j-i} = 0, \text{ при } k \leq j \leq n - 1. \quad (4.31)$$

Равенства нулю в выражении (4.31) имеют место только в том случае, если b_{ji} являются коэффициентами проверочного многочлена, определяющего кодовое слово B_i .

Вектор S с координатами S_i называется *синдромом комбинации* B_i . Координаты синдрома равны нулю для тех уравнений проверочных символов b_i , которые не искажены помехами. Отметим очень важное свойство синдрома S . Он является характеристикой ошибок, а не характеристикой отдельных кодовых слов. Ошибки, имевшие место на одинаковых местах различных кодовых слов двоичного кода, приведут к одинаковым синдромам.

4.4.3. Многотактные линейные фильтры

Операции кодирования цифровой информации циклическими кодами и их декодирования выполняются многотактными линейными фильтрами [4.2], построенными на основе регистров сдвига, путем умножения и деления с использованием порождающих или проверочных многочленов. Эти фильтры состоят из ячеек памяти (регистров сдвига), сумматоров и умножителей. Для простых полей Галуа $GF(p)$ операции сложения и перемножения коэффициентов осуществляются по $\text{mod } p$. Операции с многочленами принято выполнять, начиная с коэффициентов при старших степенях X .

Прежде чем рассматривать схемы кодеров циклических кодов, изучим как выполняются основные операции умножения и деления многочленов.

Умножение многочленов. Пусть имеем многочлен $C(X) = c_{k-1}X^{k-1} + c_{k-2}X^{k-2} + \dots + c_1X + c_0$, который надо умножить на многочлен $h(X) = h_rX^r + h_{r-1}X^{r-1} + \dots + h_1X + h_0$.

Произведение

$$C(X)h(X) = c_{k-1}h_r X^{k+r-1} + (c_{k-2}h_r + c_{k-1}h_{r-1})X^{k+r-2} + (c_{k-3}h_r + c_{k-2}h_{r-1} + c_{k-1}h_{r-2}) \times \\ \times X^{k+r-3} + \dots + (c_0h_1 + c_1h_0)X + c_0h_0.$$

Схема умножения многочленов представлена на рис. 4.6, а.

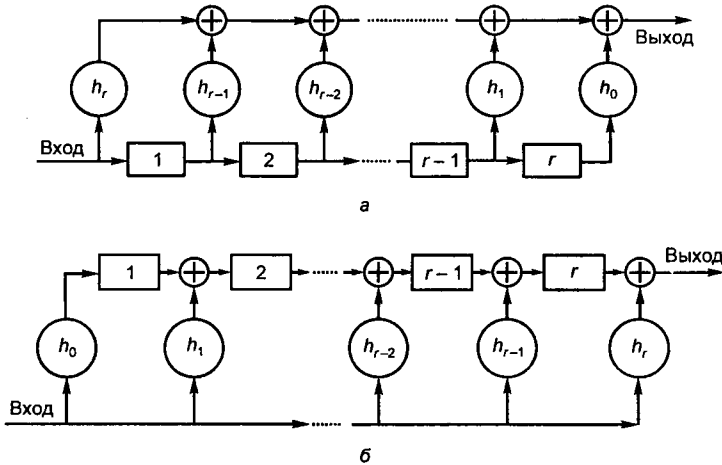


Рис. 4.6. Схема умножения многочленов:
а — вариант 1; б — вариант 2

Первоначально ячейки регистра сдвига (устройства памяти) содержат только нули. На вход схемы умножения последовательно поступают коэффициенты многочлена $C(X)$, начиная с коэффициентов при старших степенях X . Когда на вход схемы подается первый коэффициент c_{k-1} , то на выходе появляется первый коэффициент произведения $c_{k-1} h_r$. Через единицу времени (такт) на вход схемы подается второй коэффициент c_{k-2} , на выходе первой ячейки памяти имеем первый коэффициент $c_{k-2} h_r$, на выходах остальных ячеек нули. На выходе умножителя появляется сумма $c_{k-2} h_r + c_{k-1} h_{r-1}$. По аналогии через две единицы времени на вход схемы умножения многочленов подается третий коэффициент c_{k-3} , на выходе первой ячейки памяти имеем второй коэффициент $c_{k-3} h_r$, на выходе третьей — первый $c_{k-3} h_{r-1}$, на выходах остальных ячеек нули. На выходе появляется сумма $c_{k-3} h_r + \dots + c_{k-2} h_{r-1} + c_{k-1} h_{r-2}$. Дальнейшие операции производятся аналогичным образом. Так, после $r + k - 1$ сдвигов ячейки регистра сдвига содержат $0, 0, 0, \dots, 0, c_0$ и на выходе имеем $c_0 h_0$ — последний коэффициент произведения многочленов $C(X) h(X)$. Отметим, что после коэффициента c_0 на вход схемы умножения многочленов начинают подаваться нули.

Умножение многочленов можно также выполнить с помощью схемы, представленной на рис. 4.6, б.

В первый момент времени ячейки регистра сдвига содержат только нули, поэтому при поступлении на вход умножителя первого коэффициента c_{k-1} на выходе имеем произведение $c_{k-1} h_r$. После первого сдвига ячейки регистра содержат коэффициенты $c_{k-1} h_0, c_{k-1} h_1, \dots, c_{k-1} h_{r-1}$, а на выходе имеем $c_{k-1} h_{r-1} + c_{k-2} h_r$. Аналогично после второго сдвига на выходе будем иметь $c_{k-1} h_{r-2} + c_{k-2} h_{r-1} + c_{k-3} h_r$ третий коэффициент произведения многочленов и т.д.

Схемы умножения многочленов, представленные на рис. 4.6, б, могут иметь более чем один вход. Так схема с двумя входами и одним выходом представлена на рис. 4.7.

На входах схемы имеем соответственно $A_1(X)$ и $A_2(X)$, а на выходе —

$$B(X) = A_1(X) h(X) + A_2(X) f(X), \quad (4.32)$$

где

$$h(X) = h_r X^r + h_{r-1} X^{r-1} \dots + h_0;$$

$$f(X) = f_r X^r + f_{r-1} X^{r-1} \dots + f_0.$$

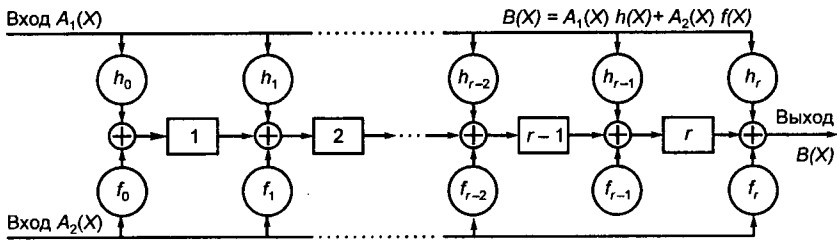


Рис. 4.7. Схема умножителя с двумя входами и одним выходом

В данном примере $h(X)$ и $f(X)$ — многочлены одинаковой степени. Если их степени не совпадают, то в качестве r можно выбрать наибольшую степень и принять старшие коэффициенты многочлена меньшей степени равными нулю.

Деление многочленов. Схема деления многочлена $A(X) = a_n X^n + a_{n-1} X^{n-1} + \dots + a_0$ на многочлен $g(X) = g_r X^r + g_{r-1} X^{r-1} + \dots + g_0$ представлена на рис. 4.8. В начальный момент состояние ячеек регистра сдвига все нули. До того момента, пока первый входной символ не достигнет конца регистра сдвига (r тактов), на выходе схемы деления будут нули. После этого на выходе появится выходной символ a_n/g_r , равный первому коэффициенту частного b_1 .

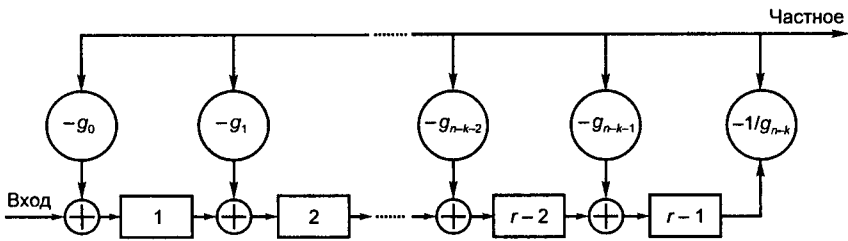


Рис. 4.8. Схема деления многочлена $A(X)$ на многочлен $g(X)$

Для получения каждого следующего коэффициента частного b_i необходимо вычесть из делимого $A(X)$ многочлен $b_i g(X)$, что осуществляется с помощью обратной связи. Через n тактов на выходе схемы появится частное от деления $A(X)$ на $g(X)$, а в ячейках регистра сдвига будет остаток. Если $A(X)$ делится на $g(X)$ без остатка, в ячейках регистра сдвига будут только нули.

Проиллюстрируем работу схемы на примере деления многочлена $A(X) = X^{11} + X^{10} + X^7 + X^4 + X^3 + X + 1$ на многочлен $g(X) = X^6 + X^5 + X^4 + X^3 + 1$ над полем $\mathbf{GF}(2)$. Схема такого делителя приведена на рис. 4.9.

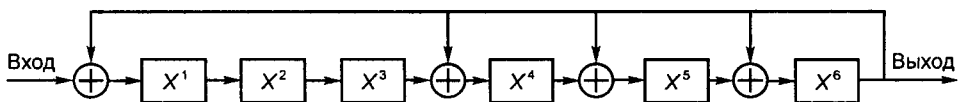


Рис. 4.9. Схема деления многочленов

Последовательные операции деления, повторяющие обычное деление столбиком, показаны в табл. 4.4.

Таблица 4.4. Пример деления многочленов

Делимое $A(X)$											Делитель $g(X)$							
X^{11}	X^{10}	0	0	X^7	0	0	X^4	X^3	0	X	1	X^6	X^5	X^4	X^3	0	0	1
X^{11}	X^{10}	X^9	X^8	0	0	X^5	0	0	0	0	0	Частное $b(X)$						
		X^9	X^8	X^7	0	X^5	X^4	X^3	0	X	1	X^5	0	X^3	0	0	1	
					X^6	X^5	X^4	0	0	X	1							
					X^6	X^5	X^4	X^3	0	0	1							
Остаток																		
								X^3	0	X	0							

Значительный интерес для реализации процедур декодирования циклических кодов представляет собой фильтр, объединяющий два фильтра. Схема такого объединенного фильтра представлена на рис. 4.10.

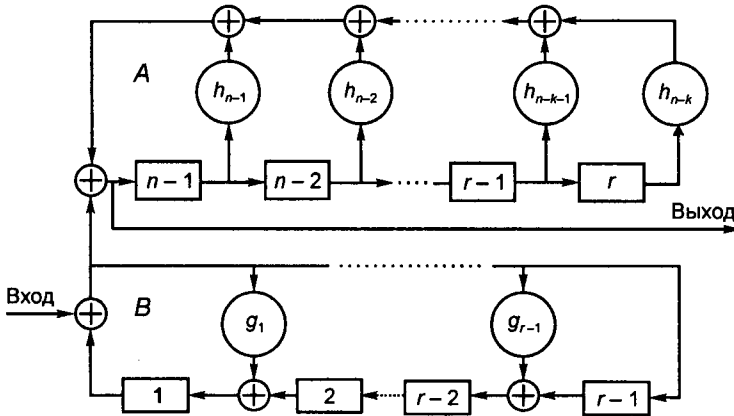


Рис. 4.10. Структура фильтра эквивалентного кольцевому регистру

Структура фильтров A и B подобрана так, что объединенный фильтр эквивалентен кольцевому регистру. Так, если на вход объединенного фильтра подать некоторую комбинацию символов $A(X)$, то через n тактов в k элементах задержки фильтра A будут записаны k последних символов комбинации $A(X)$. Фильтр B через n тактов вычислит синдром комбинации $A(X)$. На следующем такте в этом фильтре будет записан синдром циклического сдвига комбинации $A(X)$ и т.д. Аналогичный принцип заложен в основу построения описываемого ниже декодера кода Голея.

Как будет показано, приведенные в данном разделе схемы представляют собой основу кодеров, а в ряде случаев, и декодеров циклических кодов.

4.4.4. Наиболее распространенные циклические коды

Приведенные ниже циклические коды — только часть весьма обширного множества блочных кодов. Более объемное изложение блочных кодов с указанием их свойств и особенностей можно найти в [4.2–4.8].

Коды Хемминга могут быть как двоичными, так и недвоичными. Двоичные коды Хемминга это совершенные циклические коды с параметрами $n = 2^l - 1$, $k = 2^l - l - 1$ и $d = 3$, исправляющие только одиночные ошибки; l — некоторое положительное целое число. При $l = 3$ имеем известный двоичный код (7,4), порождающий многочлен которого $g(X) = X^3 + X + 1$, проверочный $h(X) = X^4 + X^2 + X + 1$. Кодер кода Хемминга (7,4), построенный на основе порождающего многочлена, представлен на рис. 4.11, а.

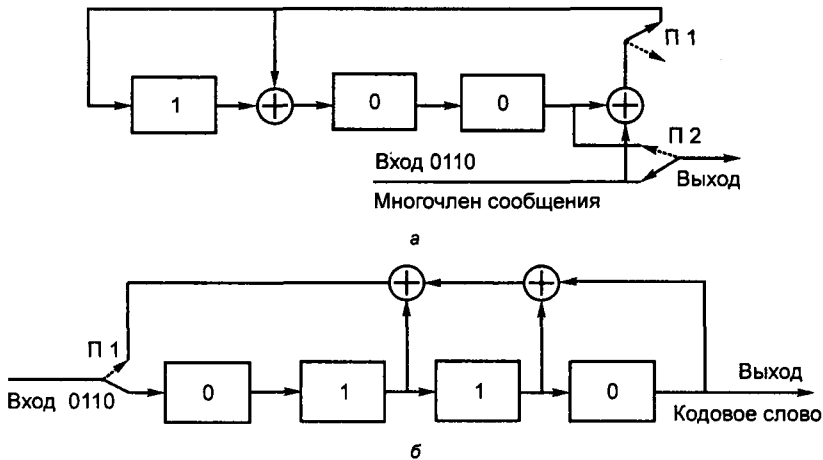


Рис. 4.11. Кодер кода Хемминга (7,4) на основе порождающего (а) и проверочного (б) многочленов

Информационная последовательность через переключатель П2 поступает на выход кодера и одновременно через переключатель П1 в верхнем (замкнутом) положении на вход кодера. Через k тактов в регистре формируются три проверочных символа, которые за три последующих такта через переключатель П2, переключенный в верхнее положение, поступают на выход кодера. Переключатель П1 синхронно с П2 переводится в нижнее (разомкнутое) положение. Таким образом, входной информационной последовательности 0110 поставлено в соответствие кодовое слово 0110001. Последовательное изменение состояний регистра иллюстрируется табл. 4.5.

Таблица 4.5. Последовательное изменение состояний регистра кодера кода Хемминга

Информация	№ шага	Содержимое регистра
	0	000
0	1	000
1	2	110
1	3	101
0	4	100

Вместо использования порождающего многочлена $g(X)$, можно построить кодер эквивалентного кода, используя проверочный многочлен $h(X) = X^4 + X^2 + X + 1$. Схема кодера

кода Хемминга (7,4), построенного на основе проверочного многочлена приведена на рис. 4.11, б. В первый момент, как и в предыдущем случае, состояние регистра все нули, переключатель П1 в нижнем положении. После четырех тактов в регистр сдвига кодера вписываются $k = 4$ информационных символа (0110), переключатель П1 переводится в верхнее положение и за семь последовательных сдвигов на выход кодера сначала поступают информационные символы (0110), затем, сформированные кодером, проверочные (0 0 1). В итоге на выходе кодера имеем кодовое слово 0 1 1 0 0 1.

Легко увидеть, что кодер, построенный по данному принципу, формирует систематический код, т.е. в начале кодового слова располагаются k информационных символов, затем $n - k$ проверочных.

Из сравнения рис. 4.11, а и б следует, что кодер на основе порождающего многочлена проще, если $k > (n/2)$, в противном случае проще будет кодер на основе проверочного многочлена.

Код Голея — двоичный совершенный циклический код с параметрами $n = 23$, $k = 12$ и $d = 7$. Его можно генерировать на основе порождающего многочлена

$$g(X) = X^{11} + X^9 + X^7 + X^6 + X^5 + X + 1. \quad (4.33)$$

Проверочный многочлен кода Голея

$$h(X) = X^{12} + X^{10} + X^7 + X^4 + X^3 + X^2 + X + 1. \quad (4.34)$$

Код гарантированно исправляет не более трех, произвольно расположенных ошибок, на любом кодовом слове. Кодер кода Голея строится по аналогии с кодерами, представленными на рис. 4.11, а и б. На рис. 4.12 показан кодер кода Голея, построенный на основе проверочного многочлена.

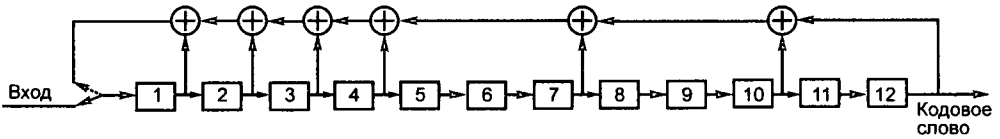


Рис. 4.12. Кодер кода Голея

В отдельных случаях удобно использовать расширенный код Голея. Его параметры $n = 24$, $k = 12$ и $d = 8$. Он получается путем добавления ко всем кодовым словам кода Голея символа проверки на четность. Расширенный код Голея уже не является совершенным.

Коды Боуза–Чоудхури–Хоквингема (БЧХ) представляют собой большой класс двоичных и недвоичных циклических кодов. На практике наиболее широко распространены двоичные коды.

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

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

$$g(X) = \text{НОК}(m_1(X) m_3(X) \dots m_{2^{m-1}}(X)), \quad (4.35)$$

причем степень $g(X)$ не превосходит mt_0 .

Коэффициенты порождающих многочленов всех двоичных БЧХ-кодов, определяемых равенством (4.35), длина которых $n \leq 1023$, приведены в табл. 9.1 работы [4.2], для $n = 7, 15, 31, 63$ и 127 — в табл. 4.2.

Многочисленные исследования двоичных циклических кодов показали, что БЧХ-коды, длина которых не превосходит 15, а также все БЧХ-коды, исправляющие двойные ошибки, являются квазисовершенными, т.е. лучшими из известных двоичными кодами. Однако при больших n эффективность их убывает, находятя многочисленные коды, исправляющая способность которых лучше, чем у БЧХ-кодов.

Кодеры двоичных БЧХ-кодов строятся по аналогии с кодерами, представленными на рис. 4.11.

Коды Рида–Соломона (РС) составляют важный подкласс недвоичных БЧХ-кодов. Символы РС-кодов выбираются из q -ичного алфавита и обозначаются $\{0, 1, 2, \dots, q-1\}$. Обычно $q = 2^k$, так что каждый q -ичный символ переносит k информационных бит. 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 $R_c = K/N$;
- количество исправляемых ошибок в кодовом слове t .

РС-код с приведенными здесь параметрами гарантированно исправляет до

$$t = \left\lfloor \frac{1}{2}(D_{\min} - 1) \right\rfloor = \left\lfloor \frac{1}{2}(N - K) \right\rfloor \quad (4.37)$$

ошибок в символах кодового слова.

Поскольку каждый символ РС-кода переносит k информационных бит, эти коды позволяют исправлять все ошибки, группирующиеся в пакки длиной не более t q -ичных символов.

Эти коды оказываются чрезвычайно востребованными для различных практических применений. *Во-первых*, они гарантированно имеют «хорошие» минимальные кодовые расстояния и, *во-вторых*, для них разработаны и производятся БИС, реализующие эффективные алгоритмы декодирования при жестких решениях, что делает возможным использование относительно длинных РС-кодов, особенно в рассматриваемых ниже каскадных конструкциях.

Мажоритарно декодируемые коды. Рассмотрим одно из правил, достаточных для того, чтобы удовлетворяющие ему коды можно было бы отнести к мажоритарно декодируемым. Таким правилом является правило ортогональных проверок. Определить это правило удобно на конкретном примере. Выберем двоичный циклический код (15, 7) с проверочным многочленом $h(X) = X^7 + X^6 + X^4 + X^0$. Этот код характеризуется минимальным кодовым расстоянием $d_{\min} = 5$ и может исправлять на длине кодового слова любые $t = 2$ ошибки. Для каждого символа произвольного кодового слова выполняется определяемая многочленом $h(X)$ проверка (см. (4.31)):

$$\begin{aligned} 0 & a_0 = a_4 + a_6 + a_7 \\ 1 & a_1 = a_5 + a_7 + a_8 \\ 2 & a_2 = a_6 + a_8 + a_9 \\ 3 & a_3 = a_7 + a_9 + a_{10} \\ 4 & a_4 = a_8 + a_{10} + a_{11} \\ 5 & a_5 = a_9 + a_{11} + a_{12} \\ 6 & a_6 = a_{10} + a_{12} + a_{13} \\ 7 & a_7 = a_{11} + a_{13} + a_{14} \\ 8 & a_8 = a_{12} + a_{14} + a_0 \\ 9 & a_9 = a_{13} + a_0 + a_1 \\ 10 & a_{10} = a_{14} + a_1 + a_2 \\ 11 & a_{11} = a_0 + a_2 + a_3 \\ 12 & a_{12} = a_1 + a_3 + a_4 \\ 13 & a_{13} = a_2 + a_4 + a_5 \\ 14 & a_{14} = a_3 + a_5 + a_6. \end{aligned} \quad (4.38)$$

Здесь сложение осуществляется по mod 2. (В поле $\text{GF}(2)$ операции сложения и умножения совпадают.)

Выберем из приведенной системы проверки, в которые обязательно входит любой один и тот же элемент, например a_0 . Он входит в проверки с номерами 0, 8, 9 и 11. Преобразовав указанные проверки так, чтобы элемент a_0 находился в левой части равенства, получим

$$\begin{aligned} 0 \quad a_0 &= a_4 + a_6 + a_7 \\ 7 \quad a_0 &= a_8 + a_{12} + a_{14} \\ 9 \quad a_0 &= a_1 + a_9 + a_{13} \\ 11 \quad a_0 &= a_2 + a_3 + a_{11} \end{aligned} \quad (4.39)$$

Характерным для приведенной здесь системы проверок является то, что ни один из элементов в правых частях всех приведенных в (4.39) проверок не повторяется дважды. Можно проверить, что это положение будет выполняться для всех символов кодовых слов данного кода. Таким образом, если для любого символа кодового слова некоторого кода ни один из символов не входит в систему проверок, построенных аналогично (4.39), два или большее число раз, то такой код представляет собой код, удовлетворяющий правилу ортогональных проверок. Это правило будет лежать в основе упрощенного алгоритма декодирования таких кодов.

4.4.5. Декодирование циклических кодов

Принципы декодирования блочных кодов на основе алгоритма максимального правдоподобия, т.е. оптимального алгоритма декодирования были представлены в 4.1 и на примерах декодирования m -ичных кодов. Такой алгоритм при всех его достоинствах в части обеспечения максимальной помехоустойчивости при больших значениях k оказывается чрезвычайно сложным в реализации. Так уже при $k = 20$ декодер должен держать в памяти более миллиона кодовых слов и, кроме того, выполнить более 10^9 вычислительных операций в реальном масштабе времени с целью определения переданного слова из k информационных бит.

Методы декодирования циклических кодов разделяют 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 ошибок более простыми путями.

Как уже указывалось выше, некоторого упрощения алгоритма декодирования для двоичных блочных кодов можно достичь, когда символы кодовых слов демодулируются независимо один от другого, т.е. посимвольно, и по каждому символу в демодуляторе принимается двоичное решение (0 или 1), т.е. осуществляется, соответствующее ДСК, жесткое декодирование.

Алгебраические алгоритмы декодирования как двоичных, так и недвоичных циклических кодов, подробно описаны, например в [4.2–4.8]. Эти алгоритмы, а также алгоритмы мягкого декодирования блочных кодов ниже рассматриваться не будут, их изложение выходит за рамки данной работы.

Среди известных алгоритмов жесткого декодирования наиболее просто реализуемым является алгоритм табличного декодирования по минимальному расстоянию. Этот алгоритм рассмотрим на примере декодирования короткого блочного двоичного кода (5, 2). Данный код образует четыре кодовых слова:

00000 01011 10101 11110.

Здесь первые два символа информационные, три последующих — проверочные. Минимальное кодовое расстояние между кодовыми словами данного кода $d_{\min} = 3$, следовательно, может быть исправлена любая, произвольно расположенная одиночная ошибка. В табл. 4.6 иллюстрируется принцип табличного декодирования короткого блокового кода. В столбцах расположим двоичные пятиразрядные комбинации, имеющие меньшее расстояние Хемминга с кодовым словом, образующим данный столбец, чем с кодовыми словами, образующими другие столбцы.

Если принята комбинация, принадлежащая первому столбцу, она отождествляется с кодовым словом 0 0 0 0 0 и декодируется как 0 0. Аналогично принимаются решения относительно комбинаций из любого другого столбца. Так, любая комбинация из третьего столбца декодируется в 1 0.

Таблица 4.6. Пример табличного декодирования

№ п/п	Кодовые слова			
	0 0 0 0 0	0 1 0 1 1	1 0 1 0 1	1 1 1 1 0
1	0 0 0 0 1	0 1 0 1 0	1 0 1 0 0	1 1 1 1 1
2	0 0 0 1 0	0 1 0 0 1	1 0 1 1 1	1 1 1 0 0
3	0 0 1 0 0	0 1 1 1 1	1 0 0 0 1	1 1 0 1 0
4	0 1 0 0 0	0 0 0 1 1	1 1 1 0 1	1 0 1 1 0
5	1 0 0 0 0	1 1 0 1 1	0 0 1 0 1	0 1 1 1 0
6	1 1 0 0 0	1 0 0 1 1	0 1 1 0 1	0 0 1 1 0
7	1 0 0 1 0	1 1 0 0 1	0 0 1 1 1	0 1 1 0 0

Наиболее просто декодер реализуется в виде запоминающего устройства, по адресам которого, соответствующим первому столбцу, записана комбинация 0 0, второму столбцу — комбинация 0 1, третьему — 1 0 и четвертому — 1 1. Современные микросхемы памяти могут позволить при относительно невысокой сложности построить табличные декодеры для кодов с $n \leq (15-20)$. Отметим, что приведенный алгоритм декодирования является алгоритмом декодирования по минимальному расстоянию Хемминга, т.е. алгоритмом максимального правдоподобия для ДСК.

Синдромное декодирование. Достаточно простой и эффективный метод жесткого декодирования циклических кодов основан на использовании синдрома. Напомним, что вектор S с координатами S_i в виде проверочных соотношений, определяемых выражениями (4.31), называется синдромом комбинации B_i .

Рассмотрим в общем виде алгоритм синдромного декодирования циклических кодов. Для всех вариантов исправляемых выбранным кодом ошибок вычисляются синдромы и составляется таблица соответствия синдромов образцам исправляемых ошибок. Для рассмотренного выше двоичного кода (5, 2) это иллюстрируется в табл. 4.7.

Таблица 4.7. Соответствие синдромов образцам исправляемых ошибок

№ п/п	Вектор ошибки	Синдром
1	0 0 0 0 0	0 0 0
2	0 0 0 0 1	0 0 1
3	0 0 0 1 0	0 1 0
4	0 0 1 0 0	1 0 0
5	0 1 0 0 0	0 1 1
6	1 0 0 0 0	1 0 1
7	1 1 0 0 0	1 1 0
8	1 0 0 1 0	1 1 1

Если код циклический, то синдром вычисляется регистром сдвига, построенным по схеме, представленной на рис. 4.8, и выполняющим деление принимаемой комбинации в виде многочлена $B(X)$ на многочлен $g(X)$. Остаток от деления $r(X)$, имеющий степень меньшую, чем $n - k$, это и есть искомым синдромом $S(X)$. Дальнейшие операции по декодированию можно выполнить, используя запоминающее устройство, адресами которого являются синдромы, а в информационной части хранятся соответствующие им векторы ошибок. Далее, сложив принятую комбинацию и вычисленный вектор ошибок, определяем наиболее вероятное переданное кодовое слово и соответствующий ему блок информационных символов.

Декодер мажоритарно декодируемых кодов. Наиболее простой алгоритм жесткого декодирования позволяют реализовать мажоритарно декодируемые коды. Как уже отмечалось, эти коды удовлетворяют требованию ортогональных проверок. Рассмотрим алгоритм декодирования данных кодов на примере декодирования двоичного циклического кода (15,7). Этот код имеет минимальное кодовое расстояние $d_{\min} = 5$ и позволяет исправлять не менее двух произвольно расположенных на кодовой комбинации ошибок. Схема такого декодера представлена на рис. 4.13. Он состоит из кольцевого регистра сдвига длины $n = 15$, четырех четырехходовых сумматоров по mod 2, мажоритарного элемента (МЭ) и двухходового сумматора по mod 2, непосредственно исправляющего ошибки. Число сумматоров определяется числом проверок в (4.39), в которые входят символы с одинаковыми показателями. В данном случае их четыре. Число входов каждого сумматора равно числу символов (слагаемых), входящих в каждую проверку в (4.39). Их также оказалось четыре. Порог МЭ, как можно будет понять из проводимого ниже анализа, выбирается равным ближайшему целому числу, которое больше половины числа четырехходовых сумматоров. В данном случае порог $P > 2$.

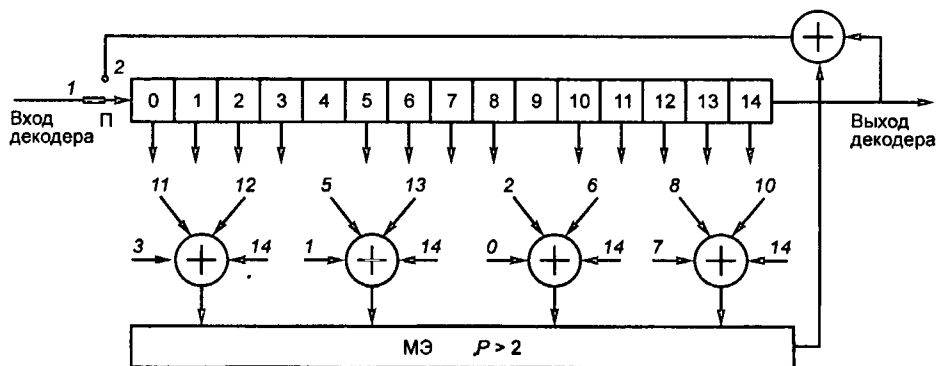


Рис. 4.13. Схема мажоритарного декодера кода (15, 7)

Вначале в кольцевой регистр вписывается принимаемая двоичная комбинация $A_i = \{a_0, a_1, \dots, a_{14}\}$. При этом переключатель Π в горизонтальном положении. Рассмотрим момент времени, когда вся комбинация вписалась в регистр, и символ a_0 находится в 14-й ячейке регистра. После этого переключатель Π ставится в положение 2. Если в принимаемой комбинации нет ошибок, то все проверки сходятся и на входах МЭ имеем только нули. На выходе МЭ имеем нуль (0), символ a_0 считается принятым правильно и инверсия символа a_0 не производится. Затем в кольцевом регистре происходит сдвиг на один такт, так что в ячейке 14 появляется символ a_1 , а в ячейке 0 — символ a_0 . Здесь также все проверки сходятся, на выходе МЭ имеем нуль (0), символ a_1 считается принятым правильно, инверсия

символа a_1 не производится и в кольцевом регистре происходит еще один сдвиг. Так проверяются все $n = 15$ символов кодового слова. После 15-го сдвига переключатель П возвращается в исходное положение, в регистр вписывается следующая комбинация и выписывается предыдущая декодированная.

Пусть в принимаемой комбинации имеется одна ошибка, например, символ a_3 принят с ошибкой. Символ a_3 входит в проверки 3, 11, 12 и 14 в (4.38), поэтому при анализе определяемых проверкой трех символов a_7, a_9 и a_{10} проверки не сойдутся, на входы МЭ вследствие ортогональности проверок будут подаваться только по одной единице, порог срабатывания $P > 2$ не будет превышен, на выходе МЭ будет ноль, символы a_7, a_9 и a_{10} будут считаться принятыми правильно и инвертироваться не будут. Аналогично будут проверены определяемые проверками 11 — символы a_0, a_2 и a_{11} , 12 — символы a_1, a_4 и a_2 и 14 — символы a_5, a_6 и a_{14} . Здесь также на входы МЭ будут подаваться по одной единице, порог МЭ не будет превышен; все указанные символы будут считаться принятыми правильно. При анализе символа принятого с ошибкой (a_3) не сойдутся все четыре проверки 3, 11, 12 и 14 в (4.38), на входы МЭ будут подаваться по единице с каждого четырехвходового сумматора, порог МЭ ($P > 2$) будет превышен, на его выходе сформируется 1, проверяемый символ (a_3) будет считаться принятым с ошибкой и при циклической перезаписи или считывании из регистра он инвертируется.

Пусть в принимаемой комбинации имеются расположенные на произвольных местах две ошибки. Здесь, если при анализе одного из принятых с ошибкой символов обе ошибки попали на символы одной проверки, данная проверка сойдется (даст 0), зато три другие проверки в силу их ортогональности будут содержать только один ошибочный символ и на соответствующих входах МЭ будет 0 1 1 1. Порог МЭ будет превышен, а ошибочный символ исправлен. Если при анализе принятого с ошибкой символа в четыре соответствующих ему проверки второй ошибочный символ не входит, на входах МЭ будет 1 1 1 1. Порог МЭ будет превышен, а ошибочный символ также исправлен.

Больше двух ошибок при произвольном их расположении данный код исправлять не должен.

Алгоритм декодирования Касами–Рудольфа. Некоторые циклические коды можно декодировать, используя алгоритм Касами–Рудольфа [4.8, 4.5]. Он позволяет получить наиболее простую реализацию декодера кода Голея. Реализацию данного алгоритма рассмотрим на примере декодирования систематического кода Голея (23,12), исправляющего любые ошибки кратности 3 и менее. Информационные символы в кодовом слове располагаются на местах с номерами от $n - k$ до $n - 1$.

Схема декодера кода Голея [4.5] представлена на рис. 4.14.

Он строится по аналогии с объединенным фильтром, схема которого представлена на рис. 4.10. Помимо двух регистров, в состав декодера входят три сумматора Σ , каждый из которых включает в свой состав и МЭ, служащий для принятия решения о необходимости исправления анализируемого символа кодовой комбинации, и элементы исправления ошибок и управления. Структура фильтров A и B подобрана так, что объединенный фильтр эквивалентен кольцевому регистру. Так, если на вход объединенного фильтра подать некоторую комбинацию символов (кодовое слово, возможно с ошибками) $A(X)$ кода Голея, то через n тактов ($n = 23$) в k элементах задержки фильтра A будут записаны k последних символов этой комбинации $A(X)$. Фильтр B через n тактов вычислит синдром S_0 комбинации $A(X)$. Напомним, синдром вычисляется в соответствии с выражением (4.31). На следующем такте в фильтре будет записан синдром S_1 первого циклического сдвига комбинации $A(X)$ и т.д.

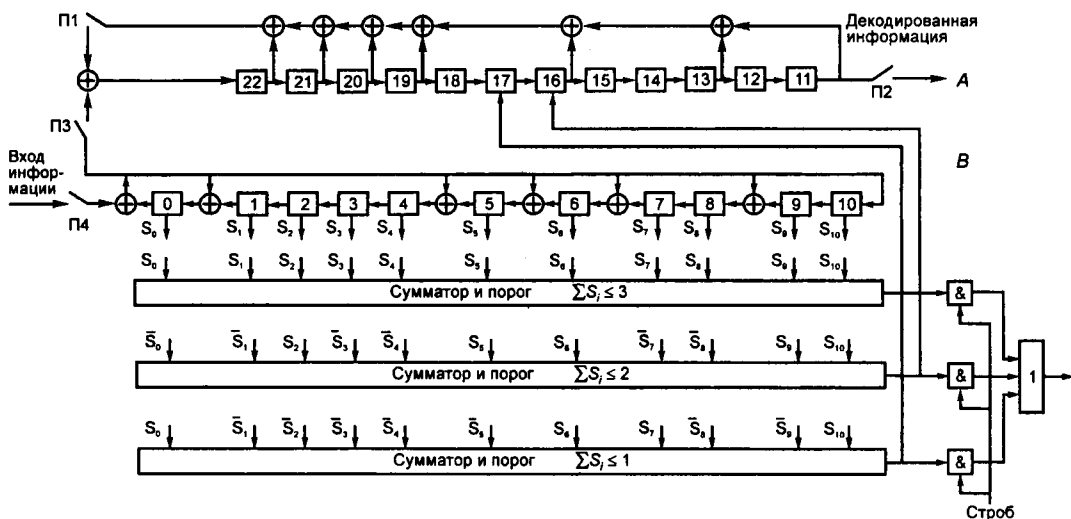


Рис. 4.14. Схема декодера кода Голея

В начале цикла декодирования в декодер вписывается принимаемая двоичная комбинация $A_i = \{a_0, a_1, \dots, a_{22}\}$. При этом переключатели П1, П3 и П4 находятся в положении «Замкнуто», а П2 — «Разомкнуто». Через 23 такта в ячейках регистра фильтра A окажутся записанными k последних символов декодируемой комбинации, а в регистре фильтра B — синдром комбинации A_i . После этого переключатель П4 переводится в положение «Разомкнуто».

Путем циклических сдвигов декодируемой комбинации $A(X)$ всегда можно добиться такого состояния, чтобы ошибки на последних k местах занимали одно из нескольких стандартных положений. Декодируемую кодовую комбинацию всегда можно циклически сдвинуть так, чтобы на последних 12 позициях находилась только одна ошибка, и она занимала либо 16-ю, либо 17-ю позицию. Число таких расстановок оказывается много меньше общего числа возможных различных расположений исправляемых ошибок. Если на информационных позициях декодируемой комбинации ошибок не произошло, то число ненулевых координат синдрома (вес синдрома ω) не превосходит числа ошибок в ней. Можно также показать, что если число ошибок в декодируемой комбинации e меньше числа исправляемых ошибок $t = 3$ и вес синдрома $\omega = e$, то на информационных позициях декодируемой комбинации ошибок нет. Если $\omega > e$, то исправляются ошибки на позициях, соответствующих первой стандартной расстановке. Вес синдрома подсчитывает сумматор Σ_1 и сравнивает его с порогом. Одновременно сумматоры Σ_2 и Σ_3 подсчитывают веса синдромов ω_1 и ω_2 , которые получились бы при коррекции на 16-й и соответственно 17-й позициях декодируемой комбинации. Это достигается путем замены на обратные тех координат синдрома, в образовании которых в соответствии с проверочными соотношениями (4.31) участвуют символы, стоящие на 16-й и 17-й позициях. На схеме декодера они отмечены черточкой над соответствующей координатой синдрома на входах сумматоров Σ_2 и Σ_3 .

Если веса $\omega_1 \leq 2$ или $\omega_2 \leq 2$, то соответствующий символ исправляется, в данном случае инвертируется. Если условия не выполняются, то на 24-м такте в объединенном фильтре производится циклический сдвиг декодируемой комбинации. После этого в фильтре B оказывается вычисленным синдром S_1 комбинации $A'(X)$, над которым выполняются операции, аналогичные описанным выше. Так декодер работает до тех пор, пока на некотором $(n + i)$ -м такте один из сумматоров не зафиксирует выполнение требуемых условий. При

этом исправляется анализируемый символ, а схема ИЛИ (1) выработает сигнал управления, переводящий переключатель ПЗ в положение «Разомкнуто». Связь фильтров A и B разрывается. В регистре фильтра A будут записаны k неискаженных символов комбинации $A^1(X)$. Далее $(n-i)$ тактов регистр фильтра A работает как кодер, после чего в нем оказываются записанными $k = 12$ неискаженных символов переданного кодового слова. Переключатель П1 переводится в положение «Разомкнуто», а П2 — в положение «Замкнуто», обратная связь в фильтре A размыкается; декодированные символы появляются на выходе декодера. Декодирование проведено за $2n$ тактов работы декодера.

Описанная схема декодера весьма проста и легко реализуется на современных микросхемах даже небольшой степени интеграции.

Приведенные примеры реализации неалгебраических алгоритмов декодирования наглядно иллюстрируют те принципы упрощения схем, которые могут быть получены при использовании определенных структурных понятий и свойств некоторых циклических кодов. Так, если декодировать код Голея, содержащий 2^{12} кодовых слов, не используя особенности его циклической структуры, потребовалось бы примерно $4096 \cdot 23$ операций перемножения, суммирования и сравнения. Использование специфические свойства кода позволило упростить схему декодера практически на три порядка.

Таблица кодов, которые можно декодировать, используя алгоритм декодирования Касами–Рудольфа, приведена в [4.5].

4.4.6. Эффективность кодирования блоковыми кодами

Эффективность помехоустойчивого кодирования удобно оценивать таким параметром, как *энергетический выигрыш от кодирования* (ЭВК). Количественно ЭВК оценивается как разность между значениями E_b/N_0 , необходимыми для обеспечения одинаковой вероятности ошибки, при использовании кодов и без кодирования при прочих равных условиях.

При анализе эффективности ограничимся только двоичными блоковыми кодами, при этом будем полагать, что прием ведется на фоне помех типа БГШ в канале без памяти, символы кода передаются противоположными сигналами с равными вероятностями, ошибки независимы, а прием оптимальный по Котельникову. При таких условиях вероятность ошибки в приеме бита информации без кодирования определяется известным выражением

$$P_{\text{ошб}} = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{2E_b/N_0}}^{\infty} e^{-x^2/2} dx = \Phi\left(\sqrt{2E_b/N_0}\right). \quad (4.40)$$

В случае кодирования блоковым кодом со скоростью $R = k/n$ вероятность ошибки в приеме кодированного символа определяется также выражением (4.40), в котором энергия бита E_b заменяется на энергию символа E_c . При этом

$$E_c = E_b R. \quad (4.41)$$

Жесткое декодирование. Напомним, что при жестком декодировании демодулятор принимает решение по каждому из принятых символов кода и на вход декодера поступают кодовые комбинации в виде двоичных символов.

Вероятность того, что кодовое слово (n, k) кода, исправляющего t ошибок, будет принято с ошибкой в первом приближении можно определить как

$$P_{\text{ошкс}} \approx \sum_{i=t+1}^n C_n^i P_{\text{ошс}}^i (1 - P_{\text{ошс}})^{n-i}. \quad (4.42)$$

Здесь $P_{\text{ошкс}}$ — вероятность ошибки в приеме кодового символа.

Строгое равенство в (4.42) имеет место только для совершенных кодов, к классу которых, как уже отмечалось выше, относятся код Голея и коды Хемминга.

Если в процессе декодирования произошла ошибка в приеме кодового слова, то в блоке из k декодированных символов часть символов будет принята с ошибками, а другая часть без них. Вероятность ошибки в приеме бита при кодировании блоковым кодом и жестком декодировании определяется выражением

$$P_{\text{ошб}} \approx \frac{1}{n} \sum_{i=1}^n i C_n^i P_{\text{ошс}}^i (1 - P_{\text{ошс}})^{n-i}. \quad (4.43)$$

Для кодов Хемминга это выражение заметно упрощается и принимает вид

$$P_{\text{ошб}} \approx P_{\text{ошс}} - P_{\text{ошс}}(1 - P_{\text{ошс}})^{n-1}.$$

Используя выражения (4.40)–(4.43), можно найти величину E_b/N_0 , необходимую для обеспечения требуемой вероятности ошибки $P_{\text{ошб}}$ при том или ином блоковом коде. Сопоставление ее со значением E_b/N_0 , вычисленным для случая отсутствия кодирования, позволяет оценить ЭВК, обеспечиваемый кодированием. На рис. 4.15 приведены характеристики помехоустойчивости некоторых блоковых кодов.

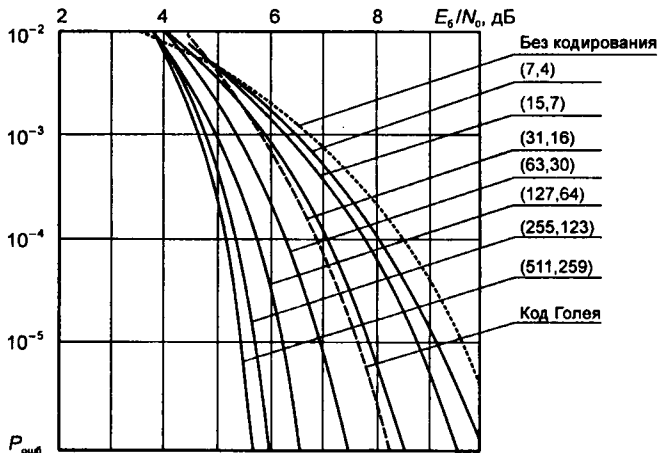


Рис. 4.15. Характеристики помехоустойчивости блоковых кодов

При оценке качества блоковых кодов большой интерес представляет выяснение тенденций изменения ЭВК от скорости кода R для разных значений n , а также от требуемой вероятности ошибки на бит. На рис. 4.16 представлены кривые зависимости ЭВК от скорости кодов R для различных длин блоковых кодов и для $P_{\text{ошб}} = 10^{-5}$.

Из анализа поведения кривых видно, что при постоянных длинах кодовых слов существует оптимальное значение скорости кода (в рассмотренных примерах это $R \approx 0,6$), при котором ЭВК максимален.

Как показывает анализ помехоустойчивости блоковых кодов с декодированием жестких решений, при $n \rightarrow \infty$ максимум ЭВК имеет место при $R = 0,5$. С увеличением длины кодового слова n при прочих равных условиях ЭВК имеет тенденцию возрастать.

Однако здесь следует отметить, что определенное влияние на ЭВК оказывают и характеристики самого кода. Так, ЭВК кода Голея (23, 12) выше ЭВК близкого к нему по значению параметров БЧХ-кода (31,16), несмотря на то, что значение n этого БЧХ-кода больше, чем для кода Голея. Скорость возрастания ЭВК с ростом n замедляется.

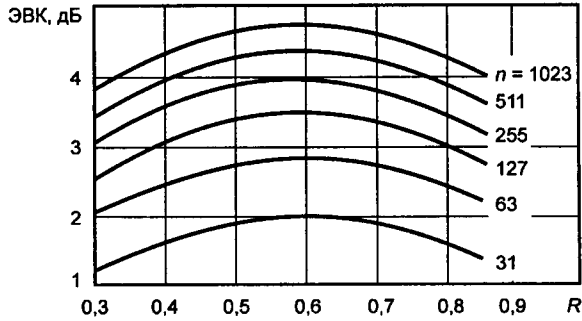


Рис. 4.16. Зависимости ЭВК от скорости кодов R для различных n

Анализ кривых, приведенных на рис. 4.15, 4.16, показывает, что при использовании блоковых кодов и жестком их декодировании, для обычно требуемых значений вероятности ошибки на бит порядка $(10^{-5}-10^{-7})$ реально достижимый ЭВК составляет (2–4) дБ. С ростом $P_{\text{ошб}}$ ЭВК уменьшается. При $P_{\text{ошб}}$ порядка 10^{-2} ЭВК близок к нулю, т.е. кодирование блоковыми кодами с жестким декодированием при низких отношениях сигнал/шум перестает быть эффективным.

Мягкое декодирование. При мягком декодировании когерентный демодулятор не принимает окончательного решения по принятым символам кода, а только вычисляет вероятность того, что принятый символ кода 1 или 0. При этом на вход декодера поступают не сами символы принимаемых кодовых комбинаций, а их коэффициенты правдоподобия (метрики) в виде трех или четырех разрядных двоичных чисел.

Из рис. 4.15, 4.16, величина ЭВК зависит и от параметров кода, и от отношения сигнал/шум (E_b/N_0). Как мы уже отмечали, декодирование с принятием по выходу демодулятора жестких решений приводит к некоторым потерям в помехоустойчивости по сравнению с мягкими решениями. (Декодирование мягких решений в литературе нередко называют приемом в целом, в то время как декодирование жестких решений — *поэлементным приемом*.)

Представляет интерес сравнить величины ЭВК, получаемые в канале с БГШ при жестком и мягком декодировании блоковых кодов. Точное выражение для вероятности ошибки в приеме кодового слова при мягком декодировании получить достаточно сложно. Поэтому многие авторы пользуются границами в оценке вероятности ошибки при декодировании мягких решений. Так, верхняя граница вероятности ошибки в приеме кодового слова [4.6] определяется следующим выражением:

$$P_M \leq \sum_{m=2}^M \left[1 - F \left(\sqrt{\frac{2E_b R_c \omega_m}{N_0}} \right) \right], \quad (4.44)$$

где ω_m — распределение весов в кодовых словах.

Для многих блоковых кодов распределение весов можно найти в [4.4].

На рис. 4.17 приведены кривые помехоустойчивости для кода Голея при декодировании жестких решений и оценка при декодировании мягких решений.

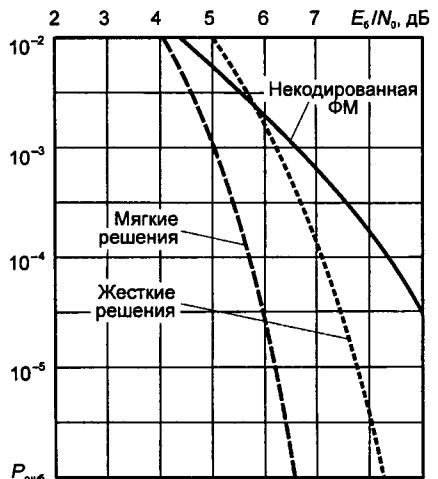


Рис. 4.17. Помехоустойчивость кода Голея: ФМ — фазовая манипуляция

Из анализа кривых видно, что при изменении вероятности ошибки от 10^{-3} до 10^{-6} мягкое декодирование в зависимости от отношения сигнал/шум позволяет получить дополнительный выигрыш от 1 дБ до 2 дБ. Исследования, проведенные в ряде работ для других блоковых кодов, показывают наличие дополнительного выигрыша от мягкого декодирования до 2,5 дБ.

Интересно найти асимптотическую оценку при $n \gg 1$ дополнительного увеличения ЭВК, получаемого в канале с БГШ при мягком декодировании двоичных блоковых кодов по сравнению с жестким. Такую оценку можно получить, сравнивая помехоустойчивость поэлементного приема и приема в целом при кодировании повторением элементарных символов. Допустим, что в канале с БГШ информация передается двумя противоположными сигналами. Кодовая последовательность каждого из этих сигналов содержит n элементарных двоичных символов [4.10]. Рассмотрим наиболее распространенный на практике случай когерентного приема.

При оптимальном поэлементном приеме вероятность того, что из n элементарных символов i будет принято с ошибкой, описывается биномиальным законом [4.11]

$$P(i) = C_n^i p^i (1-p)^{n-i}. \quad (4.45)$$

Ошибка в приеме бита будет иметь место, если более половины элементарных символов (при n — нечетном) будут приняты с ошибками, т.е.

$$P_{\text{ошб}} = \sum_{i=(n+1)/2}^n C_n^i p^i (1-p_{\text{ошс}})^{n-i}, \quad (4.46)$$

где $p_{\text{ошс}}$ — вероятность ошибки в приеме элементарного символа;

$$p_{\text{ошс}} = \Phi(2E_b/N_0 n). \quad (4.47)$$

При $n \gg 1$ выражение (4.46) достаточно сложно для вычислений. Его можно значительно упростить. Известно, что при больших n биномиальный закон (4.45) хорошо нормализуется и принимает вид

$$\omega(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right], \quad (4.48)$$

где с учетом того, что символы кодовой комбинации принимают значения 0 и 1, среднее значение случайной величины a и ее дисперсия σ^2 соответственно:

$$a = np - 0,5; \quad \sigma^2 = np(1-p). \quad (4.49)$$

Здесь $p = p_{\text{ошс}}$ (определяется выражением (4.47)).

В нашем случае при $n \gg 1$ вероятность ошибки в приеме символа $p \rightarrow 0,5$. Тогда, учитывая (4.46), (4.48) и (4.49), выражение для вероятности ошибки при поэлементном приеме бита можно получить в следующем виде:

$$P_{\text{ошб}} = \int_{-\infty}^0 \omega(x) dx = \Phi\left[\frac{\sqrt{n} \frac{2p-1}{2\sqrt{p(1-p)}}}{1}\right]. \quad (4.50)$$

Используя приближенное выражение для интеграла $\Phi(x)$ [см. 4.11] при $p \rightarrow 0,5$

$$\Phi(x) = \frac{1}{2} - \frac{1}{\sqrt{2\pi}} x \quad \text{и подставляя его в (4.50) с учетом (4.41) и (4.13), окончательно получим}$$

$$P_{\text{ошб}} = \Phi\left[\frac{\sqrt{n} \frac{2p-1}{2\sqrt{p(1-p)}}}{1}\right] = \Phi\left[\sqrt{\frac{4E_b}{\pi N_0}}\right]. \quad (4.51)$$

Сравнивая полученное выражение (4.51) с выражением для вероятности ошибки (4.40) для приема в целом, можно сделать вывод, что требуемое отношение сигнал/шум при элементарном приеме (жесткие решения) в $\pi/2$ раз, т.е. на 2 дБ выше, чем при приеме в целом (мягкие решения), т.е. ЭВК при мягком декодировании в указанных выше условиях на 2 дБ выше, чем при жестком.

Представляет интерес и нахождение асимптотической оценки ЭВК при неограниченном росте отношения сигнал/шум. Известно [4.12], что в канале с БГШ при $E_b/N_0 \rightarrow \infty$ в случае жестких и мягких решений соответственно имеем:

$$\begin{aligned} \text{ЭВК}_{\text{жестк}} &= R(t + 1); \\ \text{ЭВК}_{\text{мягк}} &= Rd. \end{aligned} \quad (4.52)$$

Из (4.52) следует, что при $E_b/N_0 \rightarrow \infty$ асимптотический выигрыш при мягких решениях в пределе в два раза (на 3 дБ) выше, чем при жестких.

4.5. Сверточные коды и методы их декодирования

4.5.1. Структура сверточных кодов

Как было отмечено в 4.2.2, большим подклассом древовидных кодов являются линейные сверточные коды. Их линейность определяется тем, что выходные символы кодера формируются как линейные комбинации входных. Сверточные коды могут быть как двоичными, так и недвоичными. Двоичные сверточные коды наиболее широко распространены в системах связи и особенно в спутниковых. Объясняется это тем, что в каналах с БГШ их использование позволяет получить весьма существенный ЭВК при относительно простой реализации декодера. Поэтому ниже будут рассматриваться двоичные сверточные коды.

Сверточные коды задаются совокупностью порождающих многочленов $g_i(X)$ и характеризуются такими параметрами, как длина кодового ограничения K , скорость $R = k_o/n_o$ и свободное расстояние $d_{св}$. Структурная схема кодера несистематического сверточного кода с параметрами $R = 1/2$, $v = 2$ и $K = 3$, заданного двумя порождающими многочленами: для верхнего сумматора $g_1(X) = 1 + X^2$, для нижнего — $G_2(X) = 1 + X + X^2$ (см. рис. 4.2). Из анализа двевидной диаграммы (см. рис. 4.3) видно, что при росте длины входной последовательности число возможных путей движения неограниченно растет, так что роль такой диаграммы сводится лишь к облегчению понимания процедуры кодирования при больших значениях длины кодового ограничения K .

Заметим, что, начиная с третьего уровня ветвей верхнее и нижнее поддерево полностью тождественны. Аналогично, на следующем уровне ветвления имеются четыре группы по четыре узла в каждой, которые также можно отождествить. Смысл такого отождествления становится более понятным при рассмотрении состояний кодера, представленного на рис. 4.2. На вход кодера последовательно поступают биты. При поступлении на вход кодера четвертого бита первый бит из регистра (кодера) выходит и далее в процессе кодирования не участвует. Следовательно, кодер, на вход которого подаются произвольные входные последовательности, например 100ху или 000ху, где крайний левый бит является самым ранним, после K (в нашем примере $K = 3$) уровней ветвления формирует одинаковые ветви. Эта повторяющаяся структура и послужила основой для построения решетчатой диаграммы [4.13], которая позволяет получить более удобное описание последовательности изменения состояний кодера. С ростом длины входной последовательности число узлов решетчатой диаграммы не растет, а остается постоянным и равным 2^{K-1} .

Для рассматриваемого сверточного кода решетчатая диаграмма приведена на рис. 4.18. При построении решетчатой диаграммы сохраним те же условные обозначения, которые были использованы при построении древовидной диаграммы, приведенной на рис. 4.3. Линии узлов решетки представляют состояния регистра: $0 = (00)$, $1 = (10)$, $2 = (01)$ и $3 = (11)$. Для рассматриваемого кода имеем четыре линии узлов. Сплошная линия перехода из одного узла диаграммы в другой соответствует ветви (двум выходным символам), генерируемой после поступления входного бита 0, штриховая — соответствует ветви, генерируемой после поступления входного бита 1.

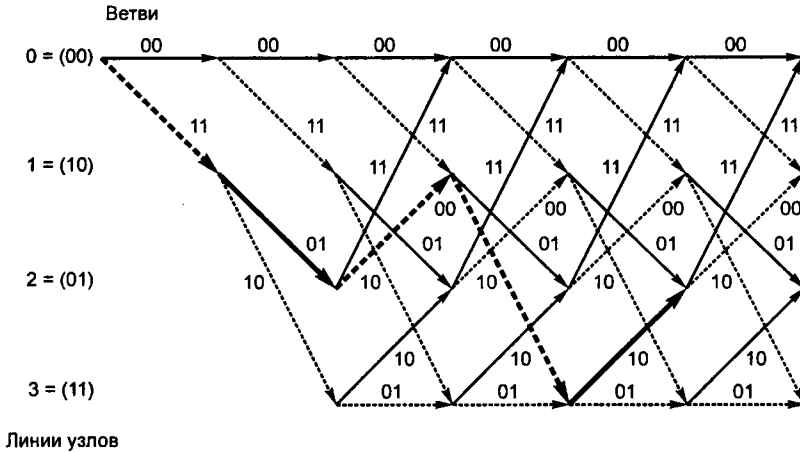


Рис. 4.18. Решетчатая диаграмма сверточного кода

Как и в структуре древовидной диаграммы, в структуре решетчатой диаграммы имеет место периодичность повторения состояния после достижения глубины ветвления K (в нашем примере $K = 3$). Так, из узла 0: при входном бите 0 происходит переход в узел 0, при этом на выход кодера проходит ветвь 00, при входном бите 1 происходит переход в узел 1; при этом на выход кодера проходит ветвь 11. Из узла 1: при входном бите 0 происходит переход в узел 2, на выход кодера проходит ветвь 01, при входном бите 1 происходит переход в узел 3, на выход кодера проходит ветвь 10. Из узла 2: при входном бите 0 происходит переход в узел 0; на выход кодера проходит ветвь 11, при входном бите 1 происходит переход в узел 1; на выход кодера проходит ветвь 00. Из узла 3: при входном бите 0 происходит переход в узел 2; на выход кодера проходит ветвь 10, при входном бите 1 происходит переход в узел 3; на выход кодера проходит ветвь 01.

Путь на решетчатой диаграмме (см. рис. 4.18), при исходном состоянии регистра 0 0 0 и входной последовательности бит 1 0 1 1 0 отмечен жирной линией. Естественно, что, как и на рис. 4.3, входной последовательности бит 1 0 1 1 0 соответствует последовательность кодированных символов 1 1 0 1 0 0 1 0 1 0. Удобство использования решетчатой диаграммы заключается в том, что из-за периодичности ее построения для «хороших» сверточных кодов при декодировании неверно выбранный в какой-либо момент времени путь с высокой вероятностью сольется с верным. В то же время, решетчатая диаграмма, как и древовидная, однозначно связывает информационную последовательность бит, последовательность состояний кодера и последовательность кодированных символов на его выходе.

Выше был рассмотрен сверточный код со скоростью $R = 1/2$. Это один из наиболее простых для рассмотрения вариантов. Более общим является рассмотрение сверточных кодов со скоростью $R = 1/n$. Здесь при поступлении на вход кодера одного бита с выхода кодера снимаются n кодированных символов. Пример такого кодера с $R = 1/3$, $K = 3$, $g_1(X) = 1$, $g_2(X) = 1 + X^2$ и $g_3(X) = 1 + X + X^2$ приведен на рис. 4.19.

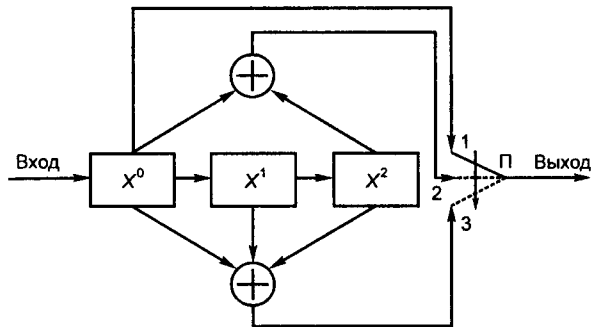


Рис. 4.19. Кодер систематического сверточного кода

Отметим, что этот код является систематическим, поскольку по цепи $g_1(X) = 1$ входные символы поступают непосредственно на выход кодера. Решетчатая диаграмма этого сверточного кода приведена на рис. 4.20.

Здесь также имеем четыре линии узлов решетки, которые представляют состояния регистра: $0 = (00)$, $1 = (10)$, $2 = (01)$ и $3 = (11)$. Однако каждая ветвь теперь содержит три символа. Отметим, что этот код систематический, поскольку входные биты непосредственно проходят на выход кодера, так как $g_1(X) = 1$.

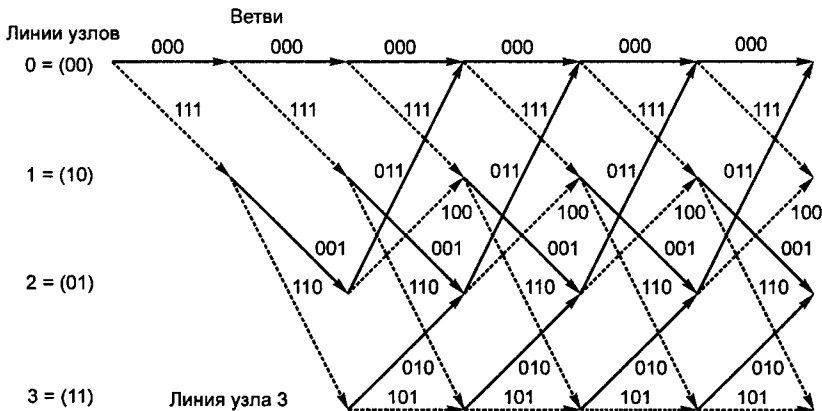


Рис. 4.20. Решетчатая диаграмма сверточного кода с $R = 1/3$

Еще более общим является рассмотрение сверточных кодов со скоростью $R = m/n$, где $m < n$. Наиболее часто такие коды строятся на основе сверточного кода с $R = 1/n$ перфорированием (удалением) из него некоторых выходных символов. Такие сверточные коды называют *перфорированными*. Известно [4.6, 4.13], что перфорированные коды имеют практически такие же характеристики помехоустойчивости, что и соответствующие им по скорости неперфорированные сверточные коды, но из-за универсальности их декодирования с декодированием кодов, из которых они образованы, такие коды оказываются более предпочтительными на практике. В качестве примера рассмотрим сверточный код с параметрами $R = 1/2$, $K = 3$, $g_1(X) = 1 + X^2$, $g_2(X) = 1 + X + X^2$ и предположим, что на выходе кодера каждый третий символ удаляется (на модулятор не подается).

Построенный таким образом кодер будет формировать для каждого двух входных бит три выходных символа, т.е. его скорость $R = 2/3$. На решетчатой диаграмме (рис. 4.21) вместо перфорированного символа 0 или 1 будет стоять X.

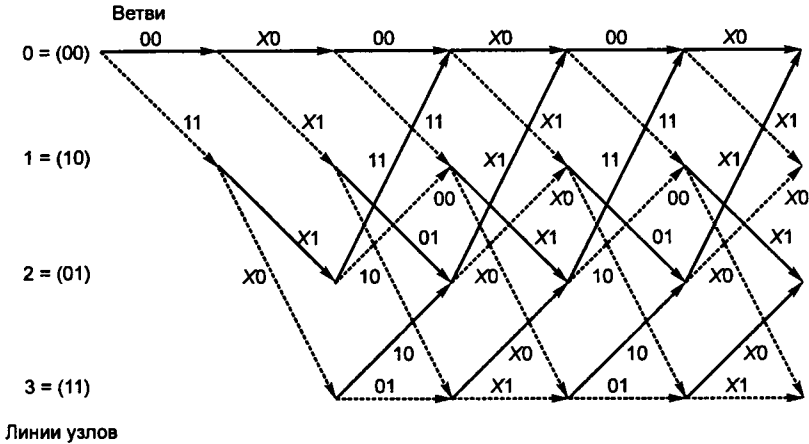


Рис. 4.21. Решетчатая диаграмма перфорированного сверточного кода

Аналогичным образом могут формироваться и другие сверточные коды со скоростями $R = m/n$.

Некоторые двоичные сверточные коды обладают свойством прозрачности. Это значит, что если значения символов на входе кодера проинвертировать, то закодированная последовательность также окажется инверсной. Порождающие многочлены таких кодов должны иметь нечетное количество слагаемых. Применение прозрачных кодов предпочтительно, когда эти коды используются в сочетании с относительной (разностной) фазовой манипуляцией (ОФМ) с присущей ей способностью принимать сигнал при возможности обратной работы, так как позволяют декодировать сигнал до устранения неопределенности знака. В этом случае применение относительного кодирования на входе кодера сверточного кода и, соответственно, относительного декодирования после декодера сверточного кода приводит к значительному выигрышу в вероятности ошибки по сравнению с вариантом относительного кодирования после кодирования информации сверточным кодом.

Свободное расстояние. Важным параметром сверточного кода (см. в 4.2), является минимальное свободное расстояние кода $d_{св}$. (В большинстве известных работ слово «минимальное» опускают.) Свободное расстояние представляет собой некоторый аналог расстояния Хемминга для блочных кодов. Код, имеющий большее свободное расстояние при прочих равных условиях, обеспечивает более высокую исправляющую способность, а следовательно, и более высокую помехоустойчивость. Поскольку сверточные коды линейны, то свободное расстояние сверточного кода $d_{св}$ равно минимальному весу наиболее короткого пути (выходной последовательности), начинающегося в нулевом состоянии и заканчивающегося также в нулевом состоянии и не попадающего в это состояние в промежуточные моменты. Для рассматриваемого кода наиболее короткий путь, определяющий его свободное расстояние, на рис. 4.22 выделен более жирной линией. Длина наиболее короткого пути 3 входных или 6 выходных символов, вес этого пути, а следовательно, и свободное расстоя-

ние рассматриваемого сверточного кода $d_{св} = 5$. Никакого другого пути длины ≤ 3 для данного кода построить невозможно.

Для сверточного кода с параметрами $R = 1/3$, $K = 3$, $g_1(X) = 1$, $g_2(X) = 1 + X^2$ и $g_3(X) = 1 + X + X^2$ длина наиболее короткого пути также 3 входных бита, а свободное расстояние $d_{св} = 6$.

Приведенный здесь метод определения свободного расстояния был использован для поиска оптимальных сверточных кодов для $K \leq 10$ и двух значений $R = 1/2$ и $R = 1/3$. Целевая функция поиска оптимальных кодов — максимум свободного расстояния. Известные оптимальные сверточные коды, заданные своими порождающими многочленами, и соответствующие им значения K и $d_{св}$, приведены в табл. 4.8.

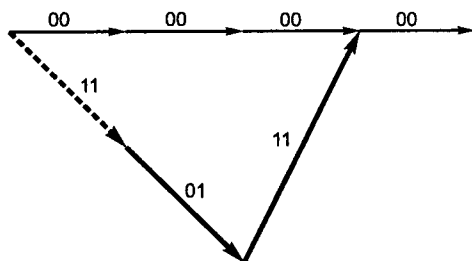


Рис. 4.22. Решетчатая диаграмма для вычисления свободного расстояния

Таблица 4.8. Оптимальные сверточные коды и их свободные расстояния

K	Порождающие многочлены	$R = 1/2$	$d_{св}$	Порождающие многочлены	$R = 1/3$	$d_{св}$
3	$g_1 = 5$ $g_2 = 7$	$R = 1/2$	5	$g_1 = 5$ $g_2 = 7$ $g_3 = 7$	$R = 1/3$	8
4	$g_1 = 15$ $g_2 = 17$	$R = 1/2$	6	$g_1 = 13$ $g_2 = 15$ $g_3 = 17$	$R = 1/3$	10
5	$g_1 = 23$ $g_2 = 35$	$R = 1/2$	7	$g_1 = 25$ $g_2 = 33$ $g_3 = 37$	$R = 1/3$	12
6	$g_1 = 53$ $g_2 = 75$	$R = 1/2$	8	$g_1 = 47$ $g_2 = 53$ $g_3 = 75$	$R = 1/3$	13
7	$g_1 = 133$ $g_2 = 171$	$R = 1/2$	10	$g_1 = 133$ $g_2 = 145$ $g_3 = 175$	$R = 1/3$	15
8	$g_1 = 247$ $g_2 = 371$	$R = 1/2$	10	$g_1 = 225$ $g_2 = 331$ $g_3 = 367$	$R = 1/3$	16
9	$g_1 = 561$ $g_2 = 753$	$R = 1/2$	12	$g_1 = 557$ $g_2 = 663$ $g_3 = 711$	$R = 1/3$	18

Примечание. Порождающие многочлены приведены в восьмеричном представлении, например, $g_1 = 23 = 10\ 011 = X^4 + X + X^0$.

При выборе сверточных кодов для практического использования следует избегать кодов с катастрофическим размножением ошибок. Для таких кодов сколь угодно длинные одинаковые последовательности кодированных символов могут порождаться различными сегментами информационных последовательностей. Ошибка в декодировании такого кода приве-

дет к ошибочному декодированию всей последующей информации, переносимой такой кодовой последовательностью. Данный эффект был подробно исследован в работе [4.15], более подробное его рассмотрение выходит за рамки книги.

Алгоритмы декодирования сверточных кодов. Известны три основных алгоритма декодирования сверточных кодов:

- пороговое декодирование [4.14];
- декодирование по алгоритму Витерби [4.15];
- последовательное декодирование [4.16].

Из числа перечисленных, алгоритмы порогового декодирования не нашли широкого применения на практике, хотя такая его модификация как многопороговое декодирование позволяет получить весьма значительный ЭВК [4.17]. Анализ этого алгоритма и его разновидностей выходит за рамки данной работы.

4.5.2. Алгоритм декодирования Витерби

Как можно было убедиться (см. рис. 4.18), 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 принятой на фоне помех последовательности кодированных символов и по нему принимать 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.

В гл. 9 будет описан алгоритм максимального правдоподобия (алгоритм Витерби) в общем виде как алгоритм демодуляции. Здесь же удобно, даже без потери общности, рассмотреть алгоритм Витерби на примере жесткого декодирования двоичного сверточного кода с параметрами $R = 1/2$, $K = 3$, $g_1(X) = 1 + X^2$, $g_2(X) = 1 + X + X^2$. Последовательность решетчатых диаграмм, иллюстрирующих процедуру декодирования, представлена на рис. 4.23, *a–e*. Пусть на вход кодера сверточного кода поступает длинная последовательность информационных бит, состоящая, например, из одних нулей. Предположим, что от демодулятора на вход декодера подаются готовые (жесткие) решения (0) или (1) по каждому из принимаемых символов. В этом случае при вычислении метрик ветвей и далее метрик путей используется метрика Хемминга.

Из-за действия помех (типа БГШ) с выхода демодулятора на вход декодера вместо кодированной последовательности 00 00 00 00...00 подается последовательность с ошибками, например, 00 10 00 10 00 00...00. Исходное состояние на решетчатой диаграмме (см. рис. 4.23, *a*) — узел 0. Из узла 0 разрешенными являются переходы в узел 0 и узел 1. При подаче с выхода демодулятора первой пары символов (00) переходу в узел 0 соответствует метрика 0, а переходу в узел 1 — метрика 2. Числа над и под узлами на рис. 4.23 соответствуют метрике пути от некоторого исходного состояния до этого узла. Из узла 1 разрешенными являются переходы в узел 2 и узел 3. На втором шаге при подаче с выхода демодулятора второй пары символов (10) переходу из узла 1 в узел 2 соответствует метрика 2, а переходу в узел 3 — метрика 0, переходу из узла 0 в узел 0 — метрика 1, а переходу в узел 1 — метрика 1. Суммируя метрики переходов (ветвей) на первых двух шагах, получим метрики четырех двухшаговых путей (см. рис. 4.23, *a*).

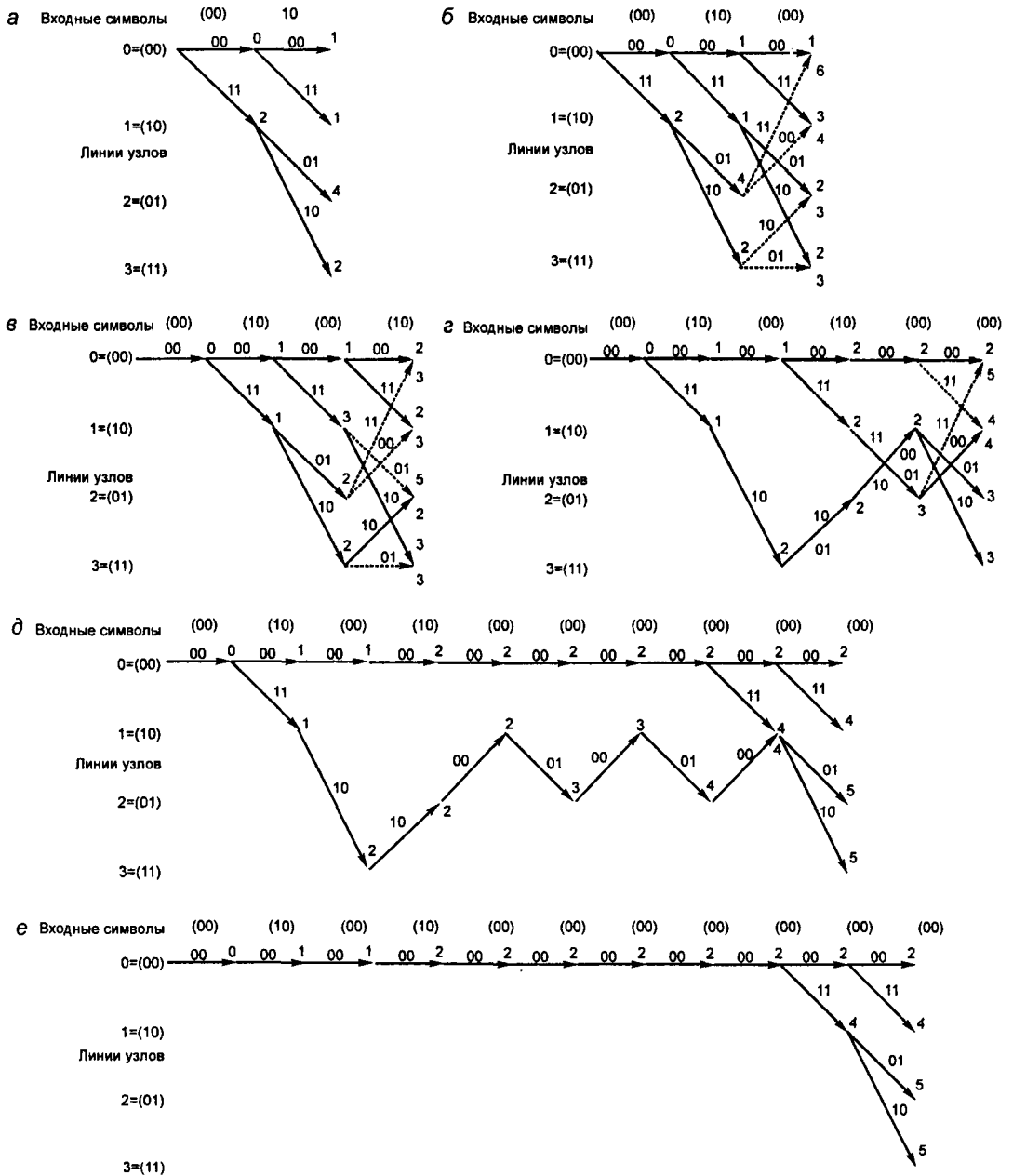


Рис. 4.23. Декодирование исправляемой комбинации ошибок

На третьем шаге процедура повторяется. С приходом третьей пары символов (00) вычисляются метрики переходов из каждого узла. На третьем шаге и далее в каждый узел приводят два пути. Например, в узел 1 приводят переходы из узла 0 с метрикой 3 и из узла 2 с метрикой 4. Из двух путей, приведших в один узел, оставляем только путь с меньшей метрикой. Это наиболее правдоподобный путь, обычно его называют выжившим (см. рис. 4.23, б). Путь с большей метрикой забывается, так как далее не используется. Все выжившие пути и их метрики хранятся в памяти декодера. В случае равенства метрик двух путей, приведших в один узел, оставляем для дальнейшего анализа один любой. Пути, не выжившие после четвертого шага, удалены. Пути, выжившие после шестого шага, представлены на рис. 4.23, г. На этом шаге уже проявляется то, что метрика «правильного» пути по линии узлов 0 оказывается наименьшей. На десятом шаге то, что метрика «правильного» пути по линии узлов 0 оказывается наименьшей, проявляется еще более явно. Здесь проявилось и то, что выжившие пути достаточно долгое время могут не совпадать. Однако на 11 шаге все выжившие до этого шага пути совпадают на первых девяти переходах.

Хотя алгоритм максимального правдоподобия требует принятия решения по выжившим путям полубесконечной глубины (длины), в данном примере можно принимать решение о первых декодированных символах (битах) после 11-го шага. Поскольку выжившие пути на первых шагах их формирования остаются на линии узлов 0, первые декодированные символы имеют значение 0, что соответствует значению переданных бит. Глубина выживших путей, на которой происходит их слияние, является случайной величиной, зависящей от вероятности символьных ошибок в канале связи. Как показали исследования [4.15], для принятия решения о значении декодированного символа достаточно иметь глубину выживших и запомненных путей $L = (5-8)K$. (Большие значения L при одинаковых K относятся к кодам с более высокими скоростями.) Переходы, по которым уже приняты решения о соответствующих им декодированных символах, из анализируемых далее путей исключаются.

Рассмотрим второй пример, когда из-за действия помех (типа БГШ) на вход декодера вместо кодированной последовательности 00 00 00 00...00 подается последовательность с ошибками, например, 11 01 00 00 00...00. Здесь на пути длиной шесть принимаемых символов появилось три ошибки, в то время как свободное расстояние рассматриваемого сверточного кода $d_{св} = 5$. Для этого случая решетчатая диаграмма приведена на рис. 4.24, а–е. Исходное состояние на решетчатой диаграмме, как и в предыдущем случае — узел 0. Отметим, что на четвертом шаге удаляется правильный (соответствующий кодируемой информации «все нули») путь. На 11 шаге все выжившие пути имеют совпадающими первые несколько переходов.

Однако, хотя при декодировании произошла ошибка, выбранный для принятия решения путь отличается от правильного на коротком отрезке, состоящем из переходов на первых трех последовательных шагах. В действительности декодированная последовательность, соответствующая выбранному пути, имеет вид 1 0 0 0 0...0, так что только один бит декодирован с ошибкой. Этот пример показывает типичную ситуацию, имеющую место при декодировании по алгоритму Витерби: ошибки влияют на сравнительно короткий участок декодированной последовательности порядка $(1-3)K$, т.е. приводят к одиночным ошибкам или коротким пакетам ошибок. Следует отметить, что алгоритм максимального правдоподобия требует принятия решения по наиболее правдоподобному пути, т.е. по пути с минимальной метрикой. Однако, так как решение о принятом бите выносится при длинах запоминаемых путей порядка $(5-8)K$, то при таких длинах с высокой вероятностью все выжившие пути исходят из одного и того же узла, поэтому практически решение о декодированном бите можно принимать по любому их выживших путей, возвращаясь по нему назад на $(5-8)K$ шагов.

Важным преимуществом алгоритма Витерби является то, что его сложность почти не увеличивается при переходе от декодирования жестких решений к мягким.

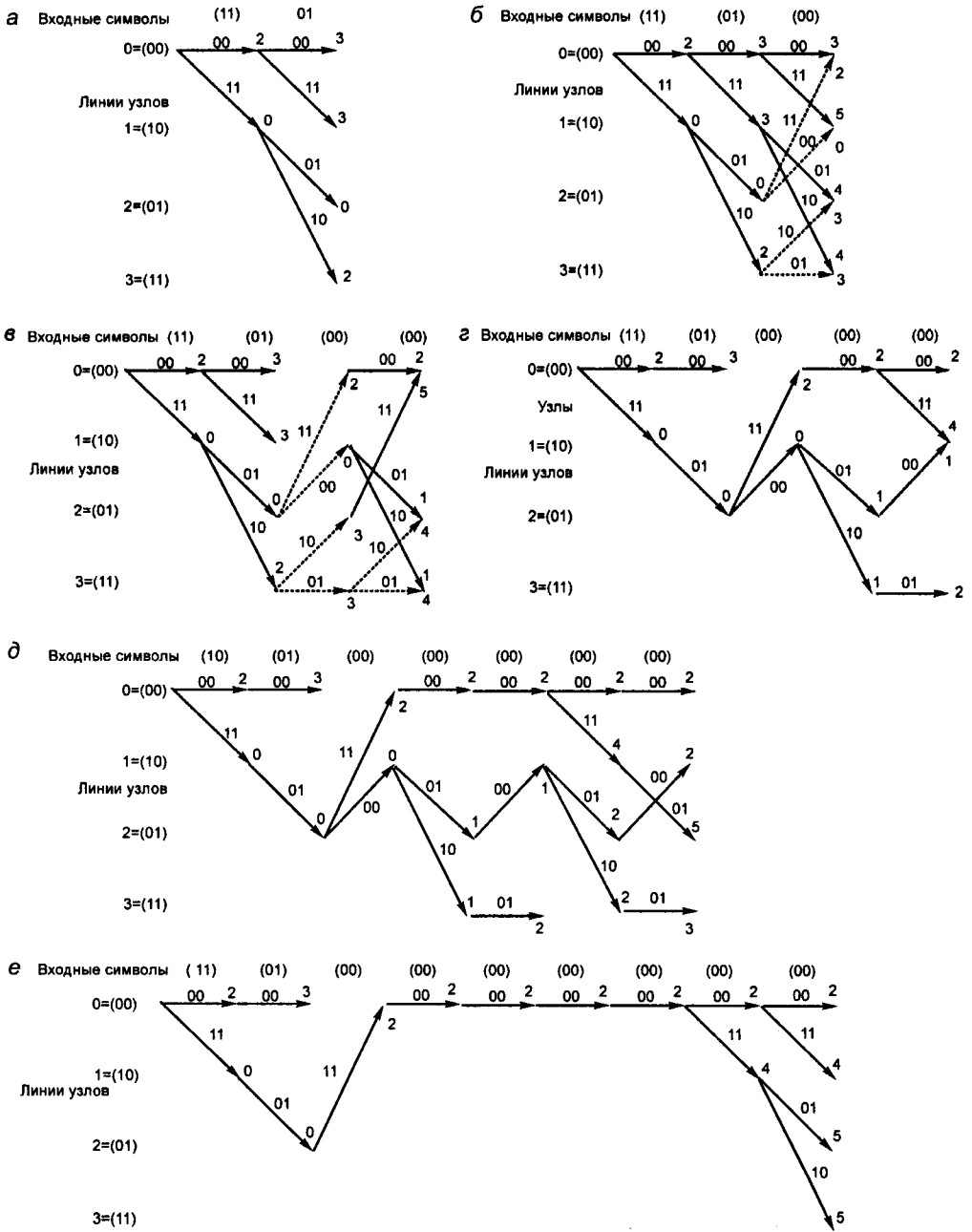


Рис. 4.24. Декодирование неисправляемой комбинации ошибок

Особенности реализации декодера Витерби. Из приведенного описания алгоритма Витерби достаточно хорошо видна его возможная реализация. Декодер должен содержать:

- устройство ветвевой синхронизации;
- устройство вычисления метрик переходов;
- устройство обновления и хранения метрик выживших путей;
- выходное решающее устройство;
- устройство управления.

Устройство синхронизации ветвей, основной задачей которого является определение момента приема первого символа ветви, постоянно оценивает скорость возрастания метрик путей. Вполне естественно, что при нормальной работе декодера ошибки появляются относительно редко и метрики выживших путей растут медленно. Если же метрики путей растут быстрее некоторого установленного предела, то это служит основанием сдвинуть фазу ожидаемого приема первого символа ветви. Этот же критерий может быть использован для определения факта обратной работы демодулятора при фазовой модуляции и в случае непрозрачных кодов.

Устройство вычисления метрик переходов при поступлении на вход декодера символов каждой новой ветви вычисляет новое множество значений метрик для каждого из разрешенных переходов.

На практике при вычислении метрик переходов неудобно пользоваться измеренными точными значениями отсчетов напряжения на выходе оптимального фильтра. Значения измеренных напряжений квантуют и представляют некоторыми числами, указывающими тот уровень квантования, которому соответствует данное напряжение. Наиболее грубое квантование на два уровня (жесткие решения), как отмечалось в разделе выше, приводит к потерям в помехоустойчивости порядка 2 дБ. Исследования, проведенные в работе [4.6], показали, что уже при восьмиуровневом квантовании при шаге квантования $0,5\sigma$, где σ — среднеквадратическое значение шума на входе решающего устройства демодулятора, потери в помехоустойчивости не превышают 0,25 дБ относительно потенциального значения отношения E_b/N_0 при аналоговых отсчетах. Квантование на 16 уровней приводит к потерям не более 0,1 дБ. При более грубом, например, четырехуровневом квантовании потери составляют примерно 1 дБ. Оптимальный шаг между соседними уровнями квантования составляет $0,5\sigma$, хотя его изменение даже в два раза не приводит к ощутимым потерям. На практике реализуется квантование на 8 или 16 уровней. Метрика принимаемого символа в этих случаях представляется соответственно трех- или четырехразрядными двоичными числами.

Проиллюстрируем процедуру вычисления метрик переходов на примере кода с параметрами $R = 1/2$, $K = 3$, $g_1(X) = 1 + X^2$, $g_2(X) = 1 + X + X^2$ при восьмиуровневом квантовании. Пусть значения отсчетов 000 и 111 являются наиболее достоверными при передаче соответственно символов 0 и 1. Предположим, что синхронизация ветвей уже установлена и что два последовательно принятые символа имеют метрики 010 (2) и 110 (6). Для ветвей (00), (01), (10) и (11) будем иметь соответственно метрики переходов

$$\begin{aligned} (00) & \text{--- } 010 + 110 = 1000, \\ (01) & \text{--- } 010 + 001 = 0011, \\ (10) & \text{--- } 101 + 110 = 1011, \\ (11) & \text{--- } 101 + 001 = 0110. \end{aligned}$$

Устройство обновления и хранения метрик выживших путей на каждом шаге декодирования суммирует метрику каждого из 2^{K-1} выживших на предыдущем шаге путей с двумя метриками разрешенных для данного кода переходов. Как видно из решетчатой диаграммы (см. рис. 4.18), на каждом шаге выживают по одному пути с минимальной метрикой из двух, приходящих в каждый узел. Здесь также сравниваются метрики двух путей, пришед-

ших в каждый узел, путь с большей метрикой забывается и далее не используется, обновленный путь с меньшей метрикой записывается в память.

Из представленного алгоритма следует, что метрики обновленных путей должны неограниченно возрастать, что влечет за собой усложнение декодера. Поэтому при реализации декодера целесообразно принять какие-либо меры, препятствующие такому возрастанию. Так, на каждом шаге, обновляя метрики выживших путей, можно добавлять не само значение метрики, а разность между реальным значением метрики и минимальной метрикой. Кроме того, можно проводить периодическую нормировку метрик выживших путей, т.е. периодически вычитать из метрик выживших путей метрику пути с минимальной метрикой. Оба указанных способа никак не снижают характеристик помехоустойчивости декодера.

Выходное решающее устройство. Каждому выжившему пути в декодере соответствует некоторая «декодированная» последовательность. Эти последовательности длиной $(5-8)K$ хранятся в 2^{K-1} регистрах. После завершения операций по определению выживших путей на каждом новом шаге регистры обмениваются своим содержимым в соответствии с выжившими путями. При этом в первую ячейку каждого регистра вписывается новый «декодированный» символ, а самый старый символ выходит из последней ячейки регистра и подается на выходное решающее устройство. На выход декодера поступит тот самый старый символ, который соответствует пути с минимальной метрикой, хотя, как отмечалось выше, для принятия решения о декодированном символе с высокой вероятностью все пути успевают слиться.

Устройство управления формирует все вспомогательные сигналы по выполнению указанных выше операций.

Сверточные коды и декодер Витерби, как уже отмечалось выше, нашли широкое применение в современных системах связи. Ряд производителей поставляет на рынок радиоэлементов БИС кодеров сверточных кодов. Среди российских предприятий следует назвать НИИМА «Прогресс», производящий БИС кодера сверточных кодов с параметрами: $g_1 = 171$, $g_2 = 133$, $K = 7$, $R = 1/2$ и $3/4$. Учитывая, что сложность реализации декодеров Витерби растет пропорционально 2^K , производители БИС кодера обычно ограничиваются созданием кодеров для $K \leq 9$.

Помехоустойчивость сверточных кодов оценивалась в ряде работ [см., например 4.4, 4.6, 4.18] на основе аддитивных границ вероятности ошибки и с помощью моделирования на ЭВМ. Используя результаты этих работ, интересно провести сравнительный анализ зависимости помехоустойчивости сверточных кодов при декодировании по алгоритму Витерби от таких параметров, как скорость кода и длина кодового ограничения. Кроме того, представляет интерес сравнение помехоустойчивости при разных методах передачи.

Характеристики помехоустойчивости при ФМ. Зависимости помехоустойчивости сверточных кодов с $K = 7$ при когерентном приеме и мягком декодировании по алгоритму Витерби для различных значений R приведены на рис. 4.25, а.

Из анализа кривых следует, что для обеспечения типового значения вероятности ошибки на бит, например, 10^{-5} необходимо иметь в зависимости от скорости кода R отношение E_b/N_0 от 5,1 до 3,7 дБ. Тем самым может быть обеспечен ЭВК по сравнению с некодированной передачей сигналов методом ФМ при когерентном приеме порядка 4,5–5,9 дБ.

Зависимости помехоустойчивости сверточных кодов для различных значений длины кодового ограничения приведены на рис. 4.25, б. Анализ зависимостей позволяет сделать вывод, что с ростом K на единицу в пределах до (8–9) для наиболее приемлемых на практике значений вероятности ошибки (10^{-3} – 10^{-6}) выигрыш от кодирования увеличивается на (0,4–0,5) дБ. Такой же характер имеют зависимости вероятности ошибки от E_b/N_0 при вариациях длины кодового ограничения при различных значениях R . При $K > 9$ сложность декодера Витерби становится фактором, ограничивающим возможность дальнейшего увеличения K .

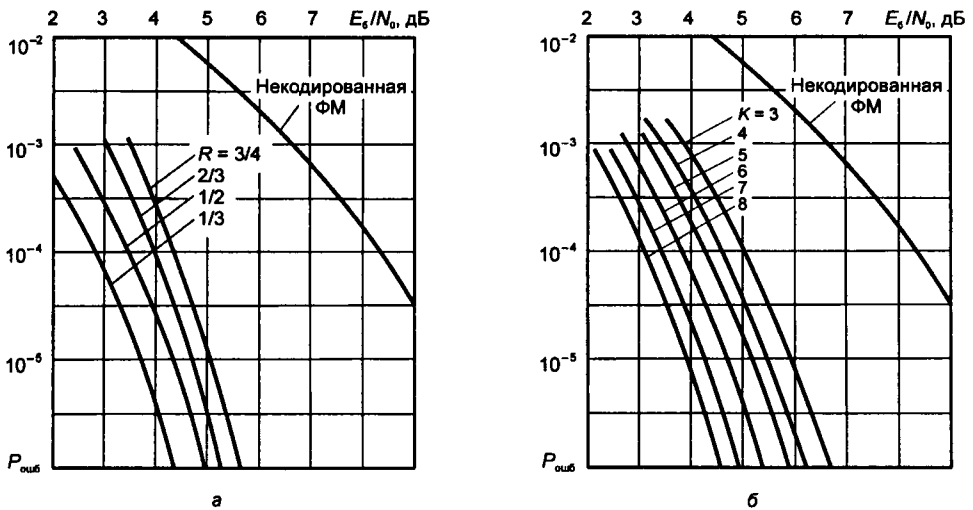


Рис. 4.25. Зависимости $P_{\text{ошб}}$ от E_b/N_0 при вариациях скорости кода (а) и длины кодового ограничения (б): ФМ — фазовая манипуляция

4.5.3. Алгоритм последовательного декодирования

Существует несколько реализаций алгоритмов последовательного декодирования сверточных кодов. Наиболее широкое применение в системах связи нашел алгоритм последовательного декодирования, предложенный Фано [4.16] еще до появления алгоритма Витерби. В отличие от декодера Витерби, этот алгоритм при приемлемой сложности реализации позволяет декодировать сверточные коды с длиной кодового ограничения $K \geq 9$. 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 реализации 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.

Несмотря на то, что из-за действия в канале шума достоверность выбора правильного пути может иногда сильно уменьшаться, общей тенденцией на значительных интервалах

времени является увеличение достоверности. Вместе с тем оценка достоверности неправильного пути иногда может возрасти, так как декодер считает этот путь как бы правильным. Алгоритм последовательного декодирования должен быстро обнаружить факт уменьшения достоверности выбранного пути и найти путь с возрастающей достоверностью. Для этого используется переменный порог Π , который в зависимости от ситуации может быть увеличен на некоторую величину Δ , называемую *приращением*. Если декодер не может двигаться вперед по выбранному пути, он пытается двигаться назад и анализирует все пути, которые не были просмотрены с текущим значением порога. В случае если ни один путь не удовлетворяет текущему порогу, его значение Π уменьшается на Δ , и декодер снова пытается продвинуться вперед. При движении вперед (Вп) декодер идет по наиболее вероятному пути, исходящему из данного узла. При движении назад (Н) осуществляется переход в узел, предшествующий данному. Движение в сторону (Вс) означает переход в узел, отличающийся от данного только последним переходом и являющийся следующим по значению метрики его пути.

В табл. 4.9 приведены правила последовательного декодирования. При малом шуме в канале декодер обычно следует правилу 1 и происходит движение вперед по правильному пути. Порог повышается, если это можно сделать так, чтобы он был на Δ меньше, чем метрика обрабатываемого пути L_k длины k шагов. Если шум становится больше и при движении вперед достоверность обрабатываемого пути оказывается меньше порога, то применяется правило 3, и декодер пытается двигаться назад, стараясь найти путь, имеющий метрику выше текущего порога. После движения назад применяется одно из правил 4 или 5 в зависимости от того, лежит L_{k-1} выше или ниже текущего порога.

Таблица 4.9. Правила последовательного декодирования

Правило	Состояние		Действие	
	Предыдущее движение	Сравнение	Порог	Движение
1	Вп или Вс	$L_{k-1} < \Pi + \Delta, L_k \geq \Pi$	Повысить (если можно)	Вп
2	Вп или Вс	$L_{k-1} \geq \Pi + \Delta, L_k < \Pi$	Без изменения	Вп
3	Вп или Вс	L_{k-1} любое, $L_k < \Pi$	Без изменения	Н
4	Н	$L_{k-1} < \Pi + \Delta, L_k$ любое	Понизить	Вп
5	Н	$L_{k-1} \geq \Pi + \Delta, L_k$ любое	Без изменения	Вс или Н

При $L_{k-1} \geq \Pi$ осуществляется переход в сторону — в следующий по достоверности узел, если такой узел существует, и движение назад, если такого узла нет. Если $L_{k-1} < \Pi$, то движение назад без сохранения порога неизменным невозможно, так что порог Π уменьшается на Δ и происходит движение вперед, и декодер пытается найти путь, оказывающийся выше уменьшенного на данном шаге порога. При этом применяется правило 2. До тех пор, пока выполнено неравенство $L_{k-1} \geq \Pi - \Delta$, порог не может быть повышен, поскольку при его повышении процедура декодирования станет бесконечной.

Рассмотрим пример последовательного декодирования простого сверточного кода с $R = 1/2$, $K = 2$, $v = 1$, кодовое дерево которого показано на рис. 4.29. Это позволит лучше понять алгоритм последовательного декодирования Фано. Предположим, что (только для упрощения описания) используются жесткие решения и переданная информационная последовательность имеет вид 1 0 0 0 0... Принятая при воздействии БГШ последовательность имеет вид S = 01 01 00 01 00 00..., т.е. в первой и четвертой ветвях имеются одиночные ошибки. Метрики путей, соответствующие выбранному коду и принятой последовательности, показаны на дереве, изображенном на рис. 4.26.

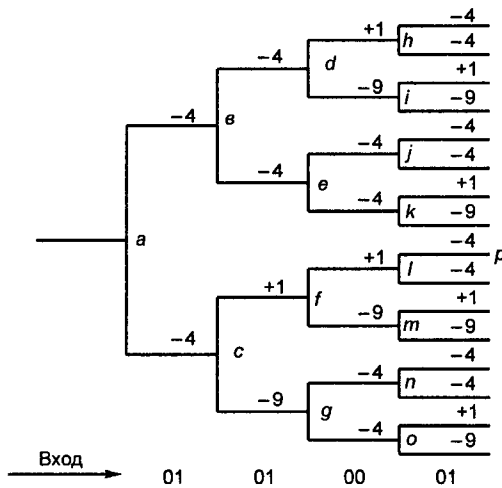


Рис. 4.26. Древоидная диаграмма последовательного декодирования

правильного решения и убывало при принятии неправильного решения. Метрика пути в узле на k -м шаге определяется выражением

$$L_k = \sum_{i=0}^k \lambda_i. \quad (4.56)$$

Такая метрика для каждого пути зависит только от данного пути и принятых символов, соответствующих первым $k+1$ узлам. Это следует из предположения о том, что декодер не имеет информации о символах, находящихся на продолжениях пути. Метрика (4.55) эквивалентна метрике максимального правдоподобия только при сравнении путей одинаковой длины. Такая метрика устроена так, что при сравнении путей различных длин более длинным путям отдается предпочтение по сравнению с более короткими, что приводит к эффективному поиску правильного пути.

В табл. 4.10 показаны шаги алгоритма Фано. Считается, что при наличии двух переходов с одинаковой метрикой (как на первом шаге) в декодере всегда вначале проверяется верхний переход. Таким образом, выполнение алгоритма начинается с попыток поиска хорошего пути в верхней половине дерева. На первом шаге вперед порог не превышает, в результате чего происходит шаг назад и уменьшение порога на -5 . Однако при движении вперед в том же направлении в узел d порог снова не превышает. Это вызывает движение назад и переход в сторону в нижнюю половину дерева для поиска лучшего пути. Движение вперед по наилучшему пути продолжается вплоть до узла p , где следующая ошибка в канале вызывает очередное непревышение порога. Затем декодер должен пройти назад по своему пути до узла a и уменьшить порог до -10 . После этого существует единственный путь, который всегда лежит над порогом.

После просмотра верхней половины дерева декодер может двигаться вперед по пути $a-c-f-l-p$, нигде не снижаясь ниже порога. Таким образом, декодером найден правильный путь. И так как в канале в нашем примере больше ошибок нет, метрика будет непрерывно возрастать только вдоль правильного пути, и декодер будет двигаться по пути только вперед.

Принято считать, что при совпадении символа перехода с принятой ветвью метрика увеличивается на $0,5$, а при несовпадении — уменьшается на $4,5$. Такой выбор метрики последовательного декодирования двоичных сверточных кодов с длиной ветви, равной n символам, предложен в работе [4.14] и позволяет декодеру выбирать наиболее достоверный путь. Это соответствует рекомендации определять метрику согласно выражению

$$\lambda_i = \sum_{j=0}^n \left\{ \log_2 \left[\frac{P(s_i^j | t_i^j)}{P(s_i^j)} \right] - B \right\}, \quad (4.55)$$

где t_i^j — переданный символ; s_i^j — принятый символ, B — параметр смещения. Обычно его выбирают так, чтобы значение метрики возрастало при принятии

Таблица 4.10. Шаги последовательного декодера при алгоритме Фано

Узел	Π	k	L_k	Окончательное значение Π	Направление движения
b	0	1	-4	0	Н
a	0	0	0	-5	Вп
b	-5	1	-4	-5	Вп
d	-5	2	-8	-5	Н
b	-5	1	-4	-5	Вс
c	-5	1	-4	-5	Вп
f	-5	2	-3	-5	Вп
l	-5	3	-2	-5	Вп
p	-5	4	-6	-5	Н
l	-5	3	-2	-5	П
f	-5	2	-3	-5	Н
c	-5	1	-4	-5	Н
a	-5	0	0	-10	Вп

Анализируя особенности реализации алгоритма последовательного декодирования, следует отметить, что последовательный декодер должен иметь буферную память для хранения входных данных для мягкого декодирования на время, в течение которого декодер ищет альтернативные пути и несколько раз возвращается назад. Когда на заданный интервал времени поиск закончен, декодер должен оперативно обработать данные по выбранному пути и освободить буферную память для начала нового поиска правильного пути на последующем участке. Иногда в течение исключительно длинных поисков буферная память может переполняться, и тогда часть информации, если не предприняты специальные меры, будет потеряна. Эта ситуация оценивается как отказ декодера от декодирования. Однако отметим, что отказ от декодирования может фиксироваться декодером и позволяет, например, в системах с переспросом запросить повторить передачу несдекодированного на данном интервале времени блока символов.

Выбор кода при последовательном декодировании оказывается не столь существенным как при других методах декодирования. Любой сверточный код можно декодировать с помощью алгоритма последовательного декодирования. Наиболее важным достоинством алгоритма последовательного декодирования является то, что сложность декодера не сильно зависит от длины кодового ограничения K . Поэтому можно увеличивать K настолько, насколько это необходимо для получения заданных характеристик помехоустойчивости.

Характеристики помехоустойчивости при последовательном декодировании. Алгоритм последовательного декодирования не является алгоритмом максимального правдоподобия и, следовательно, по помехоустойчивости при прочих равных условиях уступает алгоритму Витерби. Поэтому представляет интерес оценить реально достижимые показатели помехоустойчивости при последовательном декодировании сверточных кодов. На рис. 4.27 приведены кривые помехоустойчивости сверточных кодов с $K = 41$ и $R = 1/2$ и $1/3$ при последовательном декодировании мягких и жестких решений и сверточных кодов с $K = 9$ и $R = 1/2$ и $1/3$ при декодировании мягких решений по алгоритму Витерби.

Из анализа кривых следует, что за счет возможности декодировать коды большей сложности (с длиной кодового ограничения $K > 9$) алгоритм последовательного декодирования позволяет получить, например, для $P_{\text{ошиб}} = 10^{-5}$ выигрыш от кодирования порядка (7,5–8) дБ,

в то время как реальный выигрыш от кодирования с применением алгоритма Витерби с учетом ограничений на сложность декодера Витерби составляет (5–6) дБ.

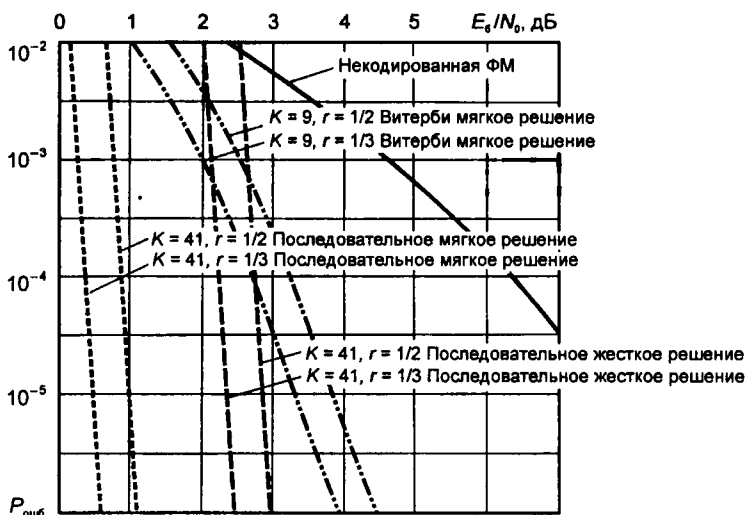


Рис. 4.27. Помехоустойчивость сверточных кодов при последовательном декодировании и декодировании по алгоритму Витерби

4.6. Каскадные коды

На практике нередко возникают задачи, когда необходимо использовать наиболее мощные коды. К числу таких задач относятся: обеспечение высокой помехоустойчивости (низкой вероятности ошибки в приеме бита) при дефиците энергетики радиолиний, что нередко имеет место в спутниковой связи; передача информации по каналам, в которых ошибки по каким-либо причинам группируются в пачки. Применение в таких случаях более длинных блочных кодов или сверточных кодов с большой длиной кодового ограничения неизбежно наталкивается на ограничения, связанные с недопустимым ростом сложности декодера. Выход из такого противоречивого положения был найден Форни [4.19], который предложил схему каскадного (многоуровневого) кодирования. На практике наиболее распространенной оказалась каскадная схема с двумя уровнями кодирования.

4.6.1. Основные принципы каскадного кодирования

Структурная схема, поясняющая основные принципы двухуровневого каскадного кодирования, приведена на рис. 4.28.

Здесь используются два кода: внешний код, обычно блочный и недвоичный, и внутренний. В качестве внешних кодов наиболее часто используются q -ичный код ($q = 2^k$) Рида–Соломона. Эти коды наиболее широко распространены, поскольку являются кодами с максимальным кодовым расстоянием для фиксированных N и K и относительно просто декодируются.

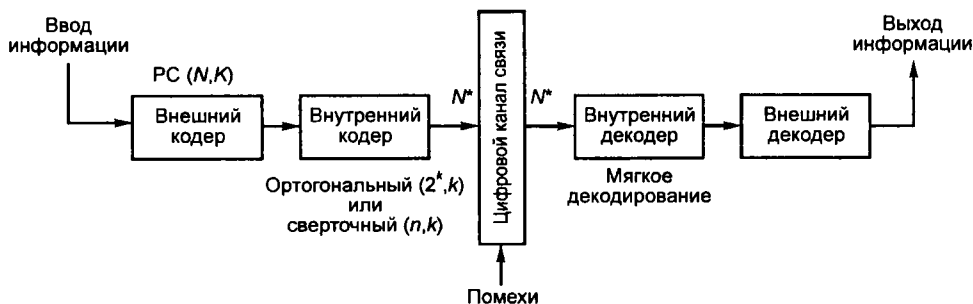


Рис. 4.28. Структурная схема каскадного кодирования

Параметры РС-кодов были определены в (4.36). Напомним их:

- длина кодового слова $N = q - 1 = 2^k - 1$, 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 $R_c = K/N$;
- количество исправляемых ошибок в кодовом слове t .

РС-код с параметрами (4.37) гарантированно исправляет до

$$t = \left\lfloor \frac{1}{2}(d_{\min} - 1) \right\rfloor = \left\lfloor \frac{1}{2}(N - K) \right\rfloor \text{ ошибок в символах кодового слова.}$$

В качестве внутренних кодов принципиально могут использоваться циклические коды, ортогональные коды и сверточные. На вопросах выбора наиболее эффективных, в смысле максимизации ЭВК, внутренних кодов для каскадных конструкций остановимся ниже.

В большинстве случаев источники цифровой информации являются двоичными. До кодирования 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 длиной N q -ичных символов или, что то же, длиной Nk двоичных символов. Далее начинается кодирование внутренним кодом.

При использовании в качестве внутреннего, относительно коротких, например, двоичных циклических кодов каждой указанной выше группе из k двоичных символов внутренним кодером ставится в соответствие кодовое слово внутреннего кода длины n двоичных символов. Длина полученного таким образом кодового слова каскадного кода $N^* = Nn$ двоичных символов, количество информационных бит $K^* = Kk$, скорость каскадного кода

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

При использовании в качестве внутренних ортогональных (в том числе биортогональных и симплексных при когерентном приеме) m -ичных кодов каждой указанной выше группе из k двоичных символов внутренним кодером ставится в соответствие кодовое слово ортогонального m -ичного кода. Поскольку, как отмечалось в 4.3, при использовании ортогональных m -ичных кодов требования к полосе пропускания канала растут пропорционально росту m , то и применение таких каскадных конструкций возможно лишь тогда, когда имеется значительный запас полосы относительно полосы, необходимой для передачи некодированной информации.

При использовании в качестве внутренних сверточных кодов поток двоичных символов кодовых слов РС-кода кодируется сверточным кодом. Длина полученного таким образом кодового слова каскадного кода $N^* = N n_0$ двоичных символов, количество информационных бит $K^* = K k_0$, скорость кода $R^* = R_{\text{внешн}} R_{\text{внутр}} = K k_0 / (N n_0)$.

Декодирование каскадных кодов осуществляется в обратном порядке. Сначала декодируется внутренний код, затем внешний. Это позволяет значительно снизить сложность декодера каскадного кода, по сравнению со сложностью декодера, который был бы необходим при одноступенчатом декодировании для обеспечения одинаковой вероятности ошибки.

4.6.2. Помехоустойчивость каскадных кодов

До последнего времени каскадные коды с внутренними циклическими кодами в системах связи использовались не часто. Объяснялось это тем, что они обеспечивали относительно невысокую помехоустойчивость, связанную с жестким декодированием внутреннего кода. Мягкое декодирование циклических кодов, обеспечивающее дополнительный ЭВК порядка 2 дБ, практически не использовалось из-за сложности реализации декодера.

Каскадные коды с внутренними ортогональными кодами (см. 4.3) применяются в системах связи более широко, особенно в тех случаях, когда канал связи широкополосный, т.е. его полоса много шире полосы, занимаемой спектром информационных сигналов и можно использовать коды с низкими скоростями кодирования ($R \ll 1$). Эти коды относительно просто декодировать в целом (т.е. мягко), поэтому их применение в каскадной конструкции достаточно эффективно. Эта эффективность сохраняется и при некогерентном приеме внутреннего кода в такой каскадной конструкции. Однако, как уже отмечалось, сложность декодера ортогональных кодов растет пропорционально 2^k , поэтому реальный интерес представляют только ортогональные коды с $k \leq 10$. На рис. 4.29, а приведены кривые помехоустойчивости каскадных кодов с внешними РС-кодами со скоростью $R = 1/2$ и внутренними ортогональными ($2^k, k$) для когерентного и некогерентного методов приема.

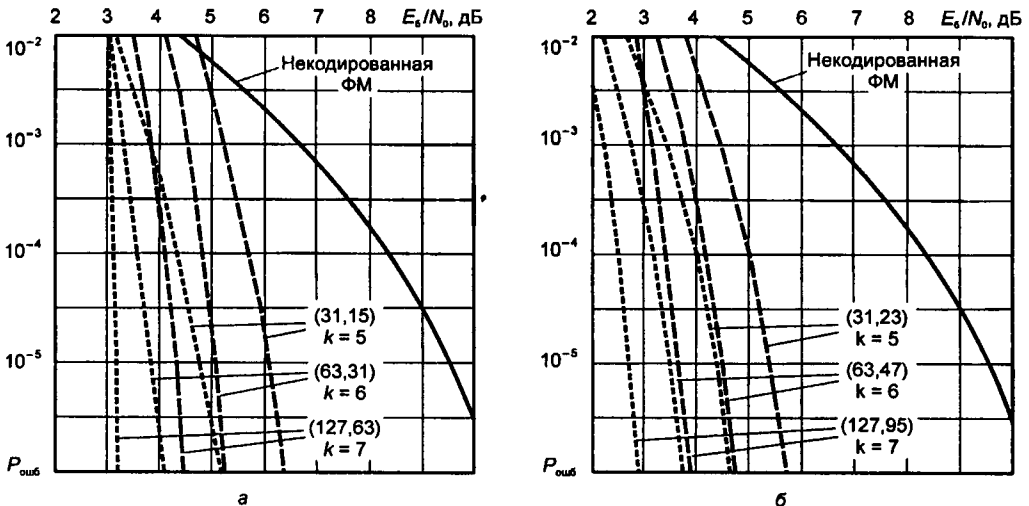


Рис. 4.29. Помехоустойчивость каскадных кодов:

а — $R = 1/2$; б — $R = 3/4$

----- — некогерентный; ————— — когерентный

На рис. 4.29, б даны аналогичные кривые помехоустойчивости с тем отличием, что здесь каскадные РС-коды имеют скорость $R = 3/4$.

Из анализа кривых:

- с ростом k при прочих равных условиях помехоустойчивость улучшается и при $k = 7$ $P_{\text{ош}} = 10^{-5}$ можно получить при $E_b/N_0 = 2,7$ дБ, так что ЭВК относительно некодированной когерентной ФМ может составить порядка 7 дБ;
- при когерентном и некогерентном приеме характеристики помехоустойчивости при $R = 3/4$ оказываются на (0,3–0,4) дБ лучше, чем при $R = 1/2$.

ЭВК при когерентном приеме каскадных кодов указанного типа выше, чем при некогерентном примерно на (0,8–1,0) дБ. Наиболее широко каскадные коды используются в конструкции с внутренними сверточными кодами. В спутниковой связи использование таких кодов из-за их высокой помехоустойчивости при относительно простой реализации диктуется действующими стандартами. В отличие от конструкции каскадных кодов с внутренними ортогональными рассматриваемые конструкции можно использовать при средних и высоких скоростях кодирования. На рис. 4.30 приведены кривые помехоустойчивости каскадных кодов с внешними РС-кодами с параметрами (255, 239) и внутренними сверточными с $K = 7$ и $R = 1/3, 1/2, 3/4$ и $7/8$.

Для оценки эффективности такого каскадного кодирования на этом же рисунке приведены кривые помехоустойчивости сверточных кодов с $K = 7$ и $R = 1/2, 3/4$ и $7/8$. Здесь декодирование сверточных кодов осуществляется по алгоритму Витерби с мягкими решениями. Из анализа кривых на рис. 4.30 видно:

- при относительно высоких скоростях кодирования ($R^* \approx 0,47$) $P_{\text{ош}} = 10^{-5}$ можно получить при $E_b/N_0 = (2,7–2,8)$ дБ, так что ЭВК при этом может составить порядка 7 дБ;
- с уменьшением R^* (за счет уменьшения скорости сверточного кода до $r = 1/3$) помехоустойчивость улучшается еще примерно на 0,5 дБ, с увеличением скорости до $R = 3/4$ ЭВК снижается примерно на 1,2 дБ, дальнейшее увеличение скорости кода до $R = 7/8$ приводит к снижению ЭВК еще примерно на 1,5 дБ.

При высоких характеристиках помехоустойчивости каскадным кодам, как, впрочем, и другим кодам, обеспечивающим высокий ЭВК, свойствен один недостаток: при относительно низких скоростях передачи информации все они дают большую задержку сигнала в радиолинии, обусловленную большими затратами времени на кодирование и декодирование.

4.7. Турбокоды и их декодирование

Разработка турбокодов (ТК) явилась наиболее заметным достижением в теории и технике помехоустойчивого кодирования за последнее десятилетие. Впервые они были описаны в 1993 г. в работе [4.20], где было продемонстрировано, что с их помощью можно практиче-

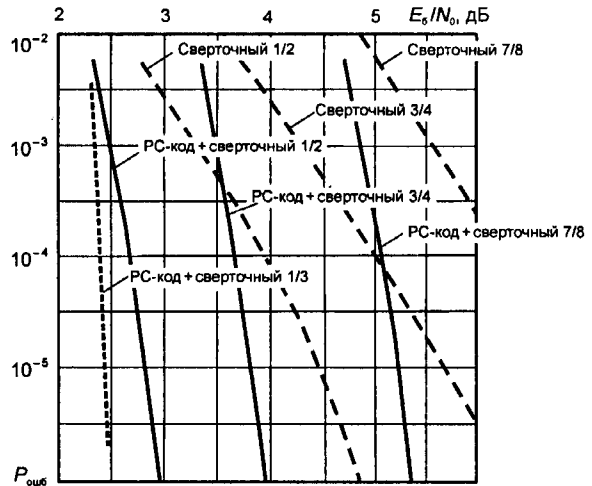


Рис. 4.30. Помехоустойчивость каскадных кодов (РС-код + сверточный)

ски вплотную приблизиться к пределу Шеннона. Так, для рассмотренных в этой работе кодов со скоростью $R = 1/2$ и предложенного итеративного алгоритма их мягкого декодирования для получения в канале связи с БГШ $P_{\text{ошб}} = 10^{-5}$ требуется отношение $E_b/N_0 = 0,7$ дБ, что на $(1,5-2)$ дБ меньше, чем удается получить при использовании лучших каскадных кодов с $R \approx 1/2$. Качественно аналогичный вывод может быть сделан и для любых других скоростей кодирования.

Разработка ТК развивается по двум направлениям:

- сверточные турбокоды (СТК), которые формируются путем параллельного соединения двух или большего числа сверточных кодеров;
- блочные турбокоды (БТК), или турбокоды произведения (Turbo Product Code, TPC), которые формируются путем последовательного соединения двух или большего числа блочных кодеров.

Как показали исследования, TPC оказываются более эффективными при относительно высоких кодовых скоростях. Рассмотрим каждое из указанных направлений более подробно.

4.7.1. Сверточные турбокоды

В 4.2.2 был рассмотрен пример (см. рис. 4.2) эффективного несистематического сверточного кода с параметрами $R = 1/2$, $K = 3$, $g_1(X) = X^2 + 1$, $g_2(X) = X^2 + X + 1$. Однако имеется

большой класс достаточно эффективных систематических кодов, среди которых можно выделить рекурсивные (с обратной связью). В качестве примера кодер такого рекурсивного сверточного кода с $K = 4$, $g_1(X) = X^3 + X^2 + 1$, $g_2(X) = X^3 + X$ представлен на рис. 4.31.

Анализ эффективности различных сверточных кодов, результаты которого приведены в работе [4.13], показал, что при относительно боль-

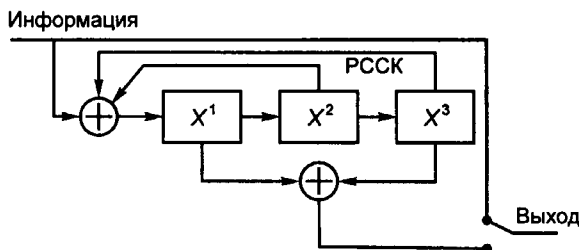


Рис. 4.31. Пример рекурсивного сверточного кодера

ших значениях E_b/N_0 и прочих равных условиях лучшие несистематические сверточные коды (НСК) более эффективны, чем систематические. Однако класс рекурсивных систематических сверточных кодов (РССК) при использовании рассмотренных ниже итеративных алгоритмов мягкого их декодирования нередко оказывается лучше самых эффективных несистематических сверточных кодов.

Кодер сверточного турбокода. Кодер эффективного СТК строится параллельным соединением двух (или нескольких) РССК с относительно небольшой длиной кодового ограничения ($K = 3-5$). Схема кодера такого турбокода приведена на рис. 4.32.

Каждый кодер РССК исходно работает со скоростью $R = 1/2$. Это означает, что на каждый информационный бит на выходе кодера РССК формируются два кодированных символа. Значение символа на систематическом выходе верхнего РССК совпадает со значением входного бита, а на втором выходе верхнего регистра формируется проверочный символ. На вход нижнего регистра кодера информационные биты подаются через перемежитель. Последний является неизбежным элементом кодера СТК, его задача внести декорреляцию во входные последовательности верхнего и нижнего регистров. Для того, чтобы осуществить перемежение, выделяется некоторый блок из k информационных бит, который предва-

рительно перед операцией кодирования записывается в матричную память, например, последовательно по строкам, а считывается из нее на вход нижнего кодера по некоторому псевдослучайному закону.

Систематический выход нижнего регистра кодера не используется, а на втором его выходе формируется проверочный бит. Таким образом, с выхода всего турбокодера на модулятор сначала поступает бит a_k с систематического выхода верхнего регистра кодера, а затем последовательно два проверочных символа: сначала с верхнего регистра x_k , затем с нижнего y_k . В результате кодовая скорость R такого турбокода в целом оказывается равной $1/3$. На основе приведенного кода с $R = 1/3$ можно сформировать перфорированный код, в котором проверочные символы с выходов двух регистров кодера снимаются по очереди (на «первом» информационном такте с верхнего, на «втором» с нижнего). В результате к первому информационному биту пары добавляется проверочный символ с верхнего регистра, а ко второму информационному биту пары с нижнего. При этом несколько снижается корректирующая способность кода, однако кодовая скорость возрастает до $1/2$. Формирование других перфорированных СТК производится так же, как описано в 4.5. Важно отметить, что составляющие ТК сверточные коды должны иметь одинаковые K и R .

Рассмотрим, какие физические принципы были заложены в представленную схему кодирования. Традиционно помехоустойчивые коды оптимизируются по критерию максимума минимального кодового (свободного) расстояния d_{\min} между кодовыми последовательностями. При этом, однако, достижение больших значений d_{\min} связано с увеличением длины кодового ограничения K и, следовательно, со значительным усложнением операции декодирования. Эффективность СТК, как следует из алгоритма их построения, определяется, в основном, не d_{\min} , а средним значением расстояний между кодовыми последовательностями $d_{\text{ср}}$, поскольку в процессе кодирования присутствует элемент случайности (перемежитель). Благодаря особенностям формирования кодовых последовательностей из двух практически независимых частей, величина $d_{\text{ср}}$ их суммы будет заметно больше, чем d_{\min} исходного сверточного кода. В то же время в отличие от сверточного кода исправляющая способность СТК будет в большей степени зависеть от распределения числа кодовых последовательностей, расположенных на расстоянии d от других кодовых последовательностей или, точнее, от вида функции распределения $S(d)$, описывающей спектр расстояний, и особенно от той ее части, где $d < d_{\text{ср}}$.

Таким образом, в отличие от других кодов, где для оценки их эффективности достаточно было знать лишь некоторые параметры этой функции, для корректной оценки эффективности СТК необходимо располагать весьма полным описанием функции распределения $S(d)$. Из структурной схемы кодера СТК ясно, что свойства функции $S(d)$ будут зависеть от длины порождающего используемый сверточный код многочлена, а также от параметров перемежителя. Исследования [4.21] показали, что при применении РССК, которые имеют неограниченную реакцию при воздействии на его вход единичного бита, достигается наибо-

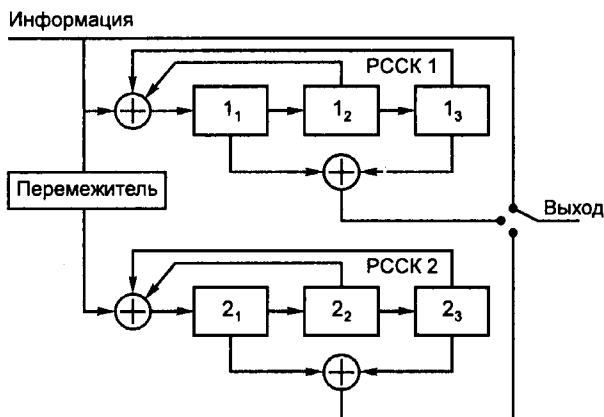


Рис. 4.32. Кодер систематического сверточного турбокода с $K = 4$

лее благоприятная форма закона распределения $S(d)$ с точки зрения его влияния на вероятность ошибочного декодирования.

Анализ многочисленных результатов экспериментальных исследований СТК, выполненных различными авторами, показал, что структура перемежителя сравнительно слабо влияет на эффективность кодирования. Те же результаты свидетельствуют о пропорциональном увеличении эффективности СТК с ростом как длины кодового ограничения сверточного кода, так и длины перемежителя.

Декодер сверточного турбокода, как это следует из построения его кодера, можно реализовать в виде двух декодеров, каждый из которых производит декодирование своего кодового блока. Укрупненная структурная схема декодера СТК приведена на рис. 4.33.

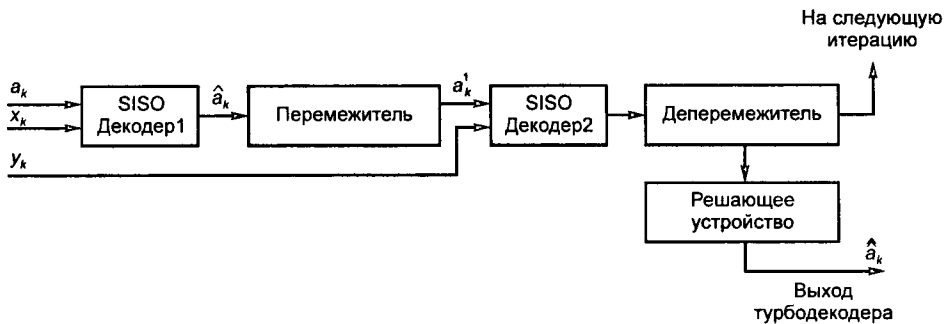


Рис. 4.33. Укрупненная структурная схема декодера СТК (одна итерация)

Кодовый блок для первого декодера представляет собой последовательность чередующихся информационных и проверочных символов a_k и x_k . Кодовый блок для второго декодера представляет собой последовательность также чередующихся символов a_k^1 и y_k , где a_k^1 — декодированная первым декодером информация, прошедшая перемежитель декодера, полностью аналогичный перемежителю кодера. Процедуру формирования декодированной первым декодером информации a_k^1 обсудим более подробно. Для формирования (мягкой) последовательности a_k^1 первым декодером можно было бы воспользоваться оптимальным по критерию максимального правдоподобия декодером Витерби. Однако алгоритм Витерби не формирует апостериорную вероятность (мягкие решения) для каждого декодированного бита. А для второго декодера СТК эта информация необходима.

Подходящий для данной задачи алгоритм был предложен в работе [4.22]. Этот алгоритм, модифицированный разработчиками СТК для учета особенностей рекурсивных кодов, позволяет формировать для каждого декодированного бита апостериорную вероятность более достоверную, чем подаваемые с демодулятора на вход первого декодера мягкие решения. В многочисленных публикациях по декодированию СТК этот модифицированный алгоритм называют *мягкий вход – мягкий выход* (Soft Input – Soft Output, SISO). Второй декодер принципиально можно было бы построить как декодер Витерби, т.е. с жесткими решениями на его выходе. Однако информация, полученная после декодирования вторым декодером, сопровождаемая сформированными им мягкими решениями a_k^2 , может значительно повысить эффективность декодера, если провести вторую итерацию декодирования, т.е. на вход первого декодера подать уже декодированные вторым декодером биты и сопровождающие их мягкие решения. Поэтому алгоритм декодирования вто-

рого декодера должен также формировать мягкие решения. Подобную операцию можно выполнять многократно. В этом и состоит принцип турбо или итеративного декодирования. Таким образом, как видно из рис. 4.33, декодер для каждой итерации декодирования СТК представляет собой каскадное соединение двух элементарных декодеров: первого и второго. Каждый из этих декодеров выносит мягкое решение о переданном символе на основе критерия максимальной апостериорной вероятности (Maximum A Posteriori, MAP), чем обеспечивается минимум вероятности ошибочного декодирования каждым элементарным декодером.

Для итеративного декодирования турбо кодов удобно использовать модульную конструкцию декодера. Укрупненная модульная схема итеративного декодера СТК приведена на рис. 4.34. В качестве модуля может быть использован декодер, представленный на рис. 4.33 без «жесткого» решающего устройства.

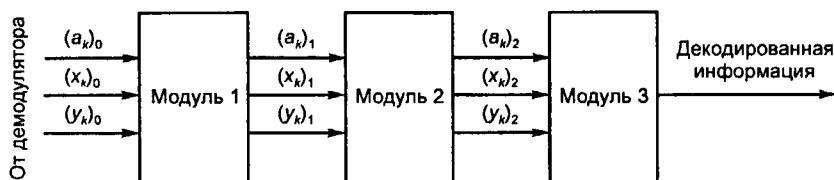


Рис. 4.34. Модульная схема итеративного декодера СТК

На второй и последующих итерациях декодирования оценка достоверности декодированных символов (бит) уточняется. Окончательное принятие решения о переданном информационном символе выносится вторым декодером на последней итерации. Изложенный алгоритм декодирования оказался чрезвычайно эффективным, поскольку каждая последующая итерация увеличивает априорную информацию о переданном символе. Окончание процесса декодирования происходит либо после выполнения заданного количества Q итерационных циклов, либо после того, как величина поправки результата декодирования достигнет установленного порога.

Оценки помехоустойчивости СТК. Основные результаты исследований помехоустойчивости СТК получены в работе [4.20]. Исследования проводились моделированием описанных процедур кодирования и итеративного декодирования методом Монте-Карло для канала с БГШ и двоичной фазовой модуляцией. Для СТК с длиной кодового ограничения $K = 5$, со скоростью $R = 1/2$ и многочленами $g_1 = 37$ и $g_2 = 21$ была вычислена зависимость вероятности ошибки на бит ($P_{\text{ошб}}$) от отношения E_b/N_0 для разного количества итераций декодирования. Для моделирования выбирались блоки длиной $N = 256 \cdot 256 = 65\,536$ информационных бит. Переключатель состоял из матричной памяти размером $256 \cdot 256$. Зависимости $P_{\text{ошб}}$ от E_b/N_0 для разных значений числа итераций декодирования Q приведены на рис. 4.35.

Анализируя представленные зависимости, следует отметить, что для любых значений $E_b/N_0 > 0,5$ дБ каждая новая итерация влечет за собой увеличение ЭВК.

В начале с ростом Q эффект от кодирования растет значительно и для $P_{\text{ошб}} = 10^{-5}$ составляет более 0,5 дБ на итерацию, для $Q > (3-4)$ рост ЭВК замедляется и при $Q > (6-8)$ становится меньше (0,1 – 0,2) дБ. Увеличение числа итераций более 18 для выбранного СТК к росту ЭВК практически не приводит. При $Q = 18$ имеем значение $P_{\text{ошб}} = 10^{-5}$ при отношении $E_b/N_0 = 0,7$ дБ, что примерно на 4 дБ лучше, чем при использовании стандартизованного сверточного кода с $K = 7$ с мягким декодированием по алгоритму Витерби.

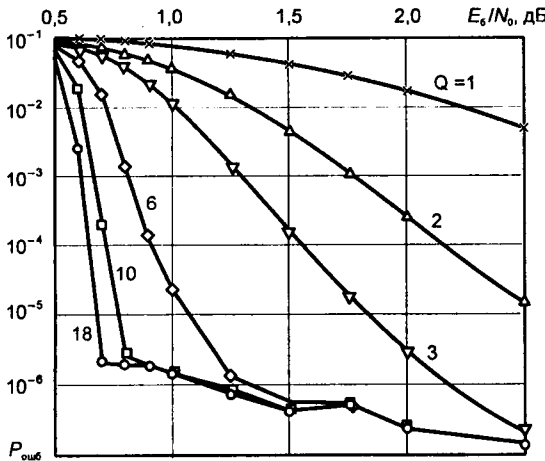


Рис. 4.35. Помехоустойчивость СТК

при $P_{ошиб} < (10^{-5} - 10^{-6})$. Исследования показали, что «точка перегиба» крутизны этих кривых сильно зависит от размера длины памяти перемежителя. Так, увеличение длины N в четыре раза практически в (4–5) раз по вероятности ошибки опускает «точку перегиба» кривых помехоустойчивости. Такое поведение ТК вполне объяснимо. Как уже отмечалось выше, исправляющая способность ТК в высокой степени зависит от вида функции распределения $S(d)$, описывающей спектр расстояний между кодовыми последовательностями, и особенно от той ее части, где $d < d_{ср}$. Вероятность появления таких ситуаций, когда расстояние d от формируемых каждым кодером кодовых последовательностей от других кодовых последовательностей близко или равно $d_{мин}$ находится в прямой зависимости от длины перемежителя N . Поэтому, когда вероятность возникновения такой ситуации становится сравнимой с требуемой вероятностью ошибки, крутизна кривых помехоустойчивости ТК резко снижается.

4.7.2. Блочные турбокоды

Блочных турбокоды (БТК) в сочетании с алгоритмом мягкого декодирования SISO с несколькими итерациями позволяют так же, как и СТК приблизиться к пределу Шеннона. Сравнительные исследования помехоустойчивости сверточных и блочных ТК показали, что БТК более эффективны, чем сверточные (при прочих равных условиях) при кодовых скоростях $R > 0,7$.

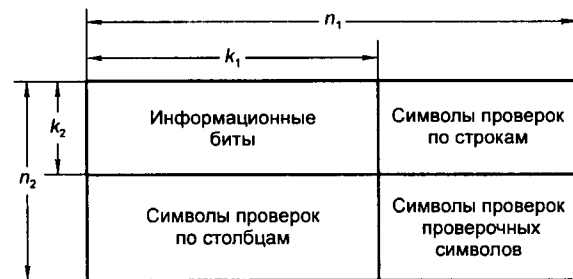


Рис. 4.36. Структура TPC, сформированного из блочных кодов

Исследование [4.20] влияния длины кодового ограничения показало, что с ростом K ($K > 5$) при $Q = (1 - 2)$ наблюдается незначительное увеличение ЭВК. Однако $P_{ошиб} = 10^{-5}$ достигается при том же значении $E_g/N_0 = 0,7$ дБ, что и при $K = 5$ и при том же числе итераций $Q = 18$. При $K = 4$ $P_{ошиб} = 10^{-5}$ достигается при значении $E_g/N_0 = 0,9$ дБ. Также при $K = 5$ исследовалось влияние различных порождающих СТК многочленов на показатели ЭВК. Моделирование показало, что лучшие результаты обеспечивают выбранные многочлены $g_1 = 37$ и $g_2 = 21$.

Особого внимания заслуживает эффект резкого снижения крутизны кривых помехоустойчивости для выбранного ТК

БТК могут формироваться как последовательным, так и параллельным соединением двух или нескольких блочных кодов [4.23]. Однако по эффективности при прочих равных условиях БТК с последовательным соединением оказываются предпочтительнее. Поэтому ниже ограничимся рассмотрением только БТК с последовательным соединением. Часто в литературе эти турбокоды называют турбокоды произведения (ТПС). Турбокод произведения можно представить в виде, приведенном на рис. 4.36.

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

Скорость ТРС определяется как $R = R_1 R_2 = k_1 k_2 / (n_1 n_2)$. Расстояние Хемминга для ТРС равно произведению расстояний составляющих кодов: $d = d_1 d_2$. В качестве исходных блочных кодов, на основе которых строятся ТРС, часто выбираются коды Хемминга (в том числе расширенные или укороченные) или коды проверки на четность. Описанный метод кодирования обеспечивает независимость и строк, и столбцов, поэтому введение более сложного перемежения перед кодированием по столбцам не приводит к какому-либо заметному положительному результату.

Символы проверок проверочных символов в отдельных турбокодеках не формируются и в цифровой канал не передаются.

В табл. 4.11 приведены порождающие многочлены для кодов Хемминга. Расширенные коды Хемминга строятся путем добавления к каждому кодовому слову символа проверки на четность.

Таблица 4.11. Порождающие многочлены для кодов Хемминга длины $7 \leq n \leq 255$

n	k	Порождающие многочлены $g(X)$
7	4	$X^3 + X + 1$
15	11	$X^4 + X + 1$
31	26	$X^5 + X^2 + 1$
63	57	$X^6 + X + 1$
127	120	$X^7 + X^3 + 1$
255	247	$X^8 + X + 1$

Основываясь на том, что с ростом отношения сигнал/шум асимптотический выигрыш от кодирования (см. (4.52)) при мягком декодировании

$$G = 10 \lg(Rd_{\min}) \text{ дБ}, \quad (4.57)$$

рассмотрим положения, которые следует учитывать при выборе «лучших» ТРС на примере кода Хемминга (63, 57). По определению коды Хемминга имеют минимальное кодовое расстояние $d = 3$. Построенный на основе двух кодов Хемминга (63, 57) ТРС имеет скорость $R = 57^2/63^2 = 0,819$ и минимальное кодовое расстояние $d = 9$. Асимптотический выигрыш от кодирования таким кодом $G = 8,67$ дБ. ТРС, построенный на базе двух расширенных кодов Хемминга (64, 57), имеет $R = 57^2/64^2 = 0,793$ и $d = 16$. Асимптотический выигрыш от кодирования таким турбокодом, как следует из выражения (4.57), составляет $G = 11,03$ дБ. Таким образом, видно, что ТРС, построенный на расширенных кодах Хемминга (64, 57), оказывается на 2,36 дБ эффективнее, чем ТРС, построенный на двух исходных кодах Хемминга. Этот вывод можно обобщить и на другие коды Хемминга.

Приведенные в табл. 4.11 коды Хемминга и построенные на их основе расширенные коды не позволяют строить ТРС для информационных блоков произвольной длины, например, с длиной $16 < k < 121$ или $121 < k < 676$ и т.д. Для согласования желаемой длины блока с длиной информационной части ТРС используют укороченные коды Хемминга.

В качестве примера рассмотрим построение ТРС для блока информационных бит длины $k = 462$ при кодовой скорости порядка 0,6. Наиболее подходящим исходным кодом является код Хемминга (31, 26). Уменьшив длину исходного внешнего кода на четыре бита, а длину внутреннего на три бита, получим код ТРС (27·28, 21·22) = (756, 462) с кодовой скоростью $R = 0,61$.

Структурная схема кодера ТРС практически не отличается от схемы кодера каскадного кода: сначала информационные биты кодируются внешним кодом. Далее кодовые слова записываются в матричную память, например, по строкам, затем считываются по столбцам и кодируются внутренним кодом. Как уже отмечалось выше, более сложное перемежение символов при кодировании заметного выигрыша в помехоустойчивости не дает.

Алгоритм декодирования блоковых турбокодов. Алгоритм итеративного декодирования блоковых турбокодов исследуем на примере простейшего ТРС кода, построенного в виде каскадного соединения двух блоковых кодов с проверкой на четность.

Структурная схема декодера ТРС с одной итерацией практически не отличается от схемы декодера каскадного кода. Однако, как и для СТК, уже на первой итерации декодер внутреннего кода по каждому декодированному символу, используя алгоритм максимума апостериорной информации, формирует мягкие решения для декодера внешнего кода. На второй итерации на вход декодера внутреннего кода подаются мягкие решения, достоверности которых уточнены после первой итерации декодирования и по строкам, и по столбцам. Окончательное решение о каждом переданном информационном бите выносится после заданного числа итераций внешним декодером.

Основным качеством, которое отличает декодер турбокода от других декодеров, является вычисление мягких решений для декодируемых символов на всех итерациях их декодирования (за исключением последней), т.е. декодирование SISO.

Рассмотрим приоритетные принципы формирования мягкого выхода (финального значения функции правдоподобия) символа a (данные) на выходе декодера $L(a^*)$.

Введем следующие обозначения:

LLR — (Log-Likelihood Ratio) — логарифм отношения правдоподобия;

$L(a)$ — априорное значения вероятности появления битов;

$L_c(x^*)$ — значение LLR до декодирования (в начале первой итерации декодирования это функция от мягких решений демодулятора);

$L_c(a^*)$ — значение LLR, вычисленное декодером в процессе декодирования.

Как показано в основополагающей работе [4.20], функции $L_c(x)$, $L_c(a^*)$ и $L(a)$ статистически независимы, поэтому финальное значение функции правдоподобия символа a на выходе каждого декодера формируется как их сумма

$$L(a^*) = L_c(x) + L_c(a^*) + L(a). \quad (4.58)$$

Вычисленное декодером значение $L(a^*)$ является вещественным числом, которое далее будет использоваться как мягкое решение для следующего декодера, а знак его принимается как жесткое решение на финальной стадии декодирования. Выражение (4.58) считается базой для построения алгоритма итеративного декодирования ТК. Для турбокодов произведений этот алгоритм будет иметь следующий вид.

1. Устанавливаются априорные значения вероятности появления битов LLR — $L(a)$. Если данные равновероятны, то $L(a) = 0$.

2. Декодируется горизонтальный код и, используя выражение (4.58), вычисляется горизонтальное значение LLR

$$L_{eh}(a^*) = L(a^*) - L_c(x) - L(a).$$

3. Для этапа вертикального декодирования принимается $L(a) = L_{eh}(a^*)$.

4. Декодируется вертикальный код и, используя выражение (4.58), вычисляется вертикальное значение LLR

$$L_{ev}(a^*) = L(a^*) - L_c(x) - L(a).$$

На этом закончилась первая итерация декодирования ТК.

На второй итерации.

5. Для стадии горизонтального декодирования принимается $L(a) = L_{ev}(a^*)$ из первой итерации.

Далее повторяются этапы 2–4.

6. После выполнения заданного количества итераций вычисляется финальное значение функции правдоподобия символа a :

$$L(a^*) = L_c(x) + L_{eh}(a^*) + L_{ev}(a^*)q.$$

7. Знак $L(a^*)$ представляет собой результат декодирования символа a .

Для вычисления значений $L_c(a^*)$ определим правила сложения двух значений LLR. Символ сложения обозначим как [+].

$$L(a_1) [+] L(a_2) = \ln \left[\frac{e^{L(a_1)} + e^{L(a_2)}}{1 + e^{L(a_1)} e^{L(a_2)}} \right];$$

$$L(a) [+] \infty = -L(a);$$

$$L(a) [+] 0 = 0.$$

(4.59)

Выражение для LLR $L_c(x_k)$ в k -й момент времени при условии, что априорные вероятности сформировать источником информации бит равный 1 или 0 одинаковы, т.е. $L(a) = 0$, будет иметь вид:

$$L_c(x_k) = \ln \left[\frac{p(x_k | a_k = +1)}{p(x_k | a_k = -1)} \right] = \ln \left\{ \frac{\frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x_k - 1}{\sigma} \right)^2 \right]}{\frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x_k + 1}{\sigma} \right)^2 \right]} \right\} =$$

$$= -\frac{1}{2} \left(\frac{x_k - 1}{\sigma} \right)^2 + \frac{1}{2} \left(\frac{x_k + 1}{\sigma} \right)^2 = \frac{2}{\sigma^2} x_k.$$

(4.60)

При $\sigma^2 = 1$

$$L_c(x_k) = 2x_k.$$

(4.61)

Алгоритм декодирования ТК проиллюстрируем на примере простейшего двумерного блочного ТК с одним символом проверки на четность [4.23]. Символы используемого кода представлены в виде табл. 4.12. Здесь a_i — входные данные; p_{ij} — символы проверки.

Таблица 4.12. Символы простейшего двумерного блочного ТК

$a_1 = 1$	$a_2 = 0$	$p_{12} = 1$
$a_3 = 0$	$a_4 = 1$	$p_{34} = 1$
$p_{13} = 1$	$p_{24} = 1$	

Проверочные символы формируются по правилу

$$p_{ij} = a_i \oplus a_j,$$

(4.62)

где знак \oplus означает сложение по mod 2. Таким образом, последовательность символов на выходе кодера ТК будет иметь вид $a_1 a_2 a_3 a_4 p_{12} p_{34} p_{13} p_{24} = 1 1 0 0 1 1 1 1 1$. На входе модулятора эта последовательность будет преобразована в биполярную $+1 +1 -1 -1 +1 +1 +1 +1$. Предположим, что помехи типа БГШ искажают переданную биполярную последовательность в принятую $\{x_i\}$, $\{x_j\} = 0,75 0,05 0,10 0,15 1,25 1,0 3,0 0,5$.

Учитывая (4.61), для закодированной таким образом последовательности символов $\{x_i\}$, $\{x_j\}$ получим следующие значения логарифма функции правдоподобия (LLR) на выходе демодулятора. Эти значения приведены в табл. 4.13.

Таблица 4.13. Значения логарифма функции правдоподобия

$L_c(x_1) = 1,5$	$L_c(x_2) = 0,1$	$L_c(x_{12}) = 2,5$
$L_c(x_3) = 0,2$	$L_c(x_4) = 0,3$	$L_c(x_{34}) = 2,0$
$L_c(x_{13}) = 6,0$	$L_c(x_{24}) = 1,0$	

Используя положения приведенного выше алгоритма декодирования для каждого символа a , можно вычислить горизонтальное значение $L_{eh}(a^*)$ при декодировании по строкам и вертикальное $L_{ev}(a^*)$ при декодировании по столбцам, а также значение $L(a^*)$ для мягкого решения после полной итерации декодирования

$$L(a^*) = L_c(x) + L_{eh}(a^*) + L_{ev}(a^*). \quad (4.63)$$

Для символа a_i кода, представленного в табл. 4.13, значение логарифма отношения правдоподобия может быть представлено в следующем виде:

$$L(a_i^*) = L_c(x_i) + L(a_i) + L_c(a_i^*) = L_c(x_i) + L(a_i) + \{[L_c(x_j) + L(a_j)] [+] L_c(x_{ij})\}, \quad (4.64)$$

где $L_c(x_j)$, $L(a_j)$ и $L_c(x_{ij})$ — каналные измерения значений LLR приема символов a_i , a_j и p_{ij} . Учитывая, что декодирование ведется последовательно по строкам и по столбцам, и используя выражения (4.63), (4.64), для каждого символа a_i рассматриваемого кода получим

$$\begin{aligned} L_{eh}(a_1^*) &= \{[L_c(x_2) + L(a_2)] [+] L_c(x_{12})\} \\ L_{ev}(a_1^*) &= \{[L_c(x_3) + L(a_3)] [+] L_c(x_{13})\} \\ L_{eh}(a_2^*) &= \{[L_c(x_1) + L(a_1)] [+] L_c(x_{12})\} \\ L_{ev}(a_2^*) &= \{[L_c(x_4) + L(a_4)] [+] L_c(x_{24})\} \\ L_{eh}(a_3^*) &= \{[L_c(x_4) + L(a_4)] [+] L_c(x_{34})\} \\ L_{ev}(a_3^*) &= \{[L_c(x_1) + L(a_1)] [+] L_c(x_{13})\} \\ L_{eh}(a_4^*) &= \{[L_c(x_3) + L(a_3)] [+] L_c(x_{34})\} \\ L_{ev}(a_4^*) &= \{[L_c(x_2) + L(a_2)] [+] L_c(x_{24})\}. \end{aligned} \quad (4.65)$$

Также учитывая, что на первой итерации декодирования по строкам $L(a_i) = 0$, поскольку априорные вероятности сформировать источником информации бит = 1 или 0 одинаковы и используя приведенные в табл. 4.13 значения логарифма функции правдоподобия $L_c(x_i)$ на выходе демодулятора, получим

$$\begin{aligned} L_{eh}(a_1^*) &= (0,1 - 0) [+] 2,5 \approx -0,1 \text{ — новое значение } L(a_1) \\ L_{eh}(a_2^*) &= (1,5 - 0) [+] 2,5 \approx -1,5 \text{ — новое значение } L(a_2) \\ L_{eh}(a_3^*) &= (0,3 - 0) [+] 2,0 \approx -0,3 \text{ — новое значение } L(a_3) \\ L_{eh}(a_4^*) &= (0,2 - 0) [+] 2,0 \approx -0,2 \text{ — новое значение } L(a_4). \end{aligned} \quad (4.66)$$

Новые значения $L(a_i)$ будут использоваться на стадии декодирования по столбцам. По аналогии с предыдущими значениями запишем

$$\begin{aligned}
 L_{ev}(a_1^*) &= (0,2 - 0,3) [+], 6,0 \approx 0,1 \text{ — новое значение } L(a_1) \\
 L_{ev}(a_2^*) &= (0,3 - 0,2) [+], 1,0 \approx -0,1 \text{ — новое значение } L(a_2) \\
 L_{ev}(a_3^*) &= (1,5 - 0,1) [+], 6,0 \approx -1,4 \text{ — новое значение } L(a_3) \\
 L_{ev}(a_4^*) &= (0,1 - 1,5) [+], 1,0 \approx 0,60 \text{ — новое значение } L(a_4).
 \end{aligned}
 \tag{4.67}$$

Новые значения $L(a_i)$ будут использоваться на следующей итерации декодирования по строкам.

Приведенные выше расчеты представляют собой первую итерацию декодирования. После ее завершения можно проследить, как на каждом из этапов меняются «мягкие решения».

На входе декодера по строкам исходно имеем (см. табл. 4.13):

$$\begin{aligned}
 L_c(x_1) &\approx 1,5 & L_c(x_2) &\approx 0,1 \\
 L_c(x_3) &\approx 0,2 & L_c(x_4) &\approx 0,3.
 \end{aligned}$$

После декодирования по строкам:

$$\begin{aligned}
 L_{eh}(a_1) &\approx -0,1 & L_{eh}(a_2) &\approx -1,5 \\
 L_{eh}(a_3) &\approx -0,3 & L_{eh}(a_4) &\approx -0,2.
 \end{aligned}$$

После декодирования по столбцам:

$$\begin{aligned}
 L_{ev}(a_1) &\approx 0,1 & L_{ev}(a_2) &\approx -0,1 \\
 L_{ev}(a_3) &\approx -1,4 & L_{ev}(a_4) &\approx 1,0.
 \end{aligned}$$

Анализируя представленные результаты изменений LLR, можно сделать вывод, что уже после завершения декодирования на первой итерации по жестким решениям (знакам $L_{ev}(a_i)$) можно получить правильный результат, хотя и с низкой достоверностью.

Проследим, как будут изменяться «мягкие решения» на второй итерации декодирования. Для вычислений воспользуемся выражениями (4.66), (4.67) и в результате получим следующие значения $L_{eh}(a_i^*)$:

$$\begin{aligned}
 L_{eh}(a_1^*) &= (0,1 - 0,1) [+], 2,5 \approx 0 \\
 L_{eh}(a_2^*) &= (1,5 - 0,1) [+], 2,5 \approx -1,6 \\
 L_{eh}(a_3^*) &= (0,3 - 1,0) [+], 2,0 \approx -1,3 \\
 L_{eh}(a_4^*) &= (0,2 - 1,4) [+], 2,0 \approx 1,2.
 \end{aligned}$$

Воспользуемся полученными значениями выражений $L_{eh}(a_i^*)$ для вычисления значений $L_{ev}(a_i)$:

$$\begin{aligned}
 L_{ev}(a_1^*) &= (0,2 - 1,3) [+], 6,0 \approx 1,1 \\
 L_{ev}(a_2^*) &= (0,3 - 1,2) [+], 1,0 \approx -1,0 \\
 L_{ev}(a_3^*) &= (1,5 - 0) [+], 6,0 \approx -1,5 \\
 L_{ev}(a_4^*) &= (0,1 - 1,6) [+], 1,0 \approx 1,0.
 \end{aligned}$$

Используя выражение (4.64) для определения $L(a_i^*)$ после второй итерации декодирования и по строкам и по столбцам, получим следующие «мягкие решения»:

$$\begin{aligned}
 L_c(x_1) &\approx 2,6 & L_c(x_2) &\approx -2,5 \\
 L_c(x_3) &\approx -2,6 & L_c(x_4) &\approx 2,5.
 \end{aligned}$$

По знакам этих решений определяем переданные биты:

$$a_1 = 1; a_2 = 0; a_3 = 0; a_4 = 1.$$

Последующие итерации декодирования могут проводиться аналогично. Однако для выбранного примера уже в результате двух итераций декодирования получены правильные решения по каждому переданному биту. На последующих итерациях достоверность принимаемых решений будет возрастать.

Глава 5

СПЕКТРАЛЬНО-ЭФФЕКТИВНАЯ ЦИФРОВАЯ МОДУЛЯЦИЯ И СИГНАЛЬНОЕ КОДИРОВАНИЕ

5.1. Спектрально-эффективные ансамбли сигналов

5.1.1. Введение

Основные (базовые) методы манипуляции приведены в 1.3. Ширина основной части спектра таких сигналов при двоичной передаче информации составляет величину порядка скорости передачи бит D . В ряде случаев такие полосы не могут быть выделены и приходится применять методы манипуляции, позволяющие использовать полосу рабочих частот более экономно. Такие методы манипуляции называются *полосно-эффективными*.

Сокращение полосы рабочих частот при неизменной скорости передачи бит может быть достигнуто за счет применения некоторых недвоичных методов цифровой модуляции. Названные методы манипуляции используют $n > 2$ элементарных сигналов (ЭС). Каждый из них может иметь любую форму, однако типовыми являются ЭС, построенные из отрезков гармонических радиосигналов с постоянной для каждого такого элемента амплитудой, частотой и фазой. В простейшем случае ЭС представляет собой один отрезок такого сигнала, причем различные возможные ЭС отличаются амплитудами (АМ), частотами (ЧМ) или фазами (ФМ) радиосигналов. В более сложном случае каждый ЭС представляет собой определенную последовательность отрезков гармонических колебаний с определенным законом чередования их параметров. Такой вариант используется, например, при построении фазоманипулированных широкополосных сигналов (ФМ ШПС), описанных в гл. 7.

Другой вариант выбора элементарных сигналов основан на одновременной (параллельной) передаче определенных наборов отрезков гармонических колебаний; различным элементарным сигналам соответствуют различные такие наборы. Можно ожидать, что максимальная спектральная эффективность достигается при рациональном использовании именно простейшего из перечисленных вариантов, поэтому ниже только он и рассматривается. Заметим, что отмеченный выше вариант сигналов с параллельной структурой часто можно свести к следующей схеме: поток подлежащих передаче двоичных символов сообщения делится на некоторое число l потоков с соответственно уменьшенной скоростью следования символов в каждом; для передачи этих потоков организуется l параллельно функционирующих каналов связи, но уже с простейшими сигналами. Такой вариант передачи, оказывается иногда весьма полезным в многолучевых каналах (более подробно см. гл. 9). В рассматриваемых здесь каналах с БГШ он не дает ничего нового по сравнению с методами манипуляции, использующими простейшие элементарные сигналы.

Итак, в соответствии с вышеизложенным, будем полагать, что каждый ЭС представляет собой отрезок гармонического колебания некоторой частоты ω длительностью T_s . Тогда

можно утверждать, что ширина основного спектра посылки составляет $\Delta f = 1/T_3$ и лежит около частоты ω . В той же полосе в квадратуре к данной, т.е. со сдвигом фазы на 90° , можно расположить еще одну ортогональную ей посылку. Больше двух ортогональных посылок этой длительности в данной полосе не существует. Поэтому в качестве геометрического образа указанной полосы частот можно рассматривать плоскость с заданной на ней прямоугольной системой координат, на которой в соответствие ортогональным элементарным сигналам поставлены векторы, ориентированные вдоль осей координат. Условимся длины векторов принимать равными \sqrt{E} , где E — энергия соответствующей посылки. Таким образом, полоса частот $\Delta f = 1/T_3$ образует пространство размерности 2 относительно ЭС длительностью T_3 . Можно искусственно ограничить размерность данного пространства единицей, если договориться рассматривать сигналы, образованные только одной из квадратур. Что касается увеличения размерности пространства элементарных посылок, то оно возможно только при расширении полосы. Так, полоса $\Delta f = \nu/T_3$, где ν — целое больше или равное единице число, позволяет сформировать пространство размерности $m = 2\nu$ для ЭС длительностью T_3 . Таким образом, в этой полосе может быть создан ансамбль из m попарно ортогональных сигналов.

Как это следует из гл. 4, такие сигналы обеспечивают достаточно высокую энергетическую эффективность передачи. Там же было показано, что еще большей энергетической эффективностью обладает ансамбль биортогональных сигналов, который образуется путем добавления к ортогональным сигналам еще и всех им противоположных. В частности, для $m = 4$ биортогональный ансамбль, используемый при ФМ4, при неизменной длительности посылки, обеспечивает двукратный энергетический выигрыш по сравнению со случаем использования ортогонального ансамбля, эквивалентного ЧМ. Следует отметить, что с ростом m энергетический выигрыш от использования биортогонального ансамбля вместо ортогонального сокращается и, в пределе, при $m \rightarrow \infty$, вообще исчезает. Последнее следует из того, установленного в гл. 4 факта, что использование ортогонального ансамбля в пределе реализует пропускную способность канала связи и, следовательно, не может проигрывать биортогональному кодированию.

Пусть теперь некоторый метод манипуляции использует ЭС длительностью T_3 , так что число таких сигналов, передаваемых за единицу времени, составляет $B = 1/T_3$; эту величину мы будем именовать *символьной скоростью*. Если при данном методе манипуляции применяется ансамбль из m ЭС, т.е. на каждом тактовом интервале длиной T_3 выбирается один из m возможных сигналов, то информация, транслируемая за данный интервал, составляет $k = \log_2 m$ бит. Поэтому скорость передачи бит, именуемая битовой скоростью, составляет $D = kB$. Если при этом используется сигнальное пространство размерности $n \geq m/2$, то можно выбирать ЭС из энергетически эффективных биортогональных ансамблей. Если соответствующей полосой рабочих частот данная система связи не располагает, приходится использовать сигнальные пространства меньшей размерности чем $m/2$, а следовательно, и энергетически менее эффективные множества ЭС. Последнее неизбежно приведет к снижению помехоустойчивости.

При сравнении различных сигналов необходимо иметь количественные оценки их энергетической и спектральной эффективности. Показатель энергетической эффективности — нормированное отношение сигнал/шум h_6 — был введен в 3.2. Остановимся на показателе спектральной эффективности. В общем, он определяется отношением скорости передачи сообщений, обеспечиваемых сигналами данного класса, к занимаемой ими полосе частот. Такого рода показатель γ уже вводился в 2.4. При этом скорость передачи определялась так, как это установлено в теории информации, т.е. с учетом статистики сообщений источника.

В настоящей и следующих главах по тем же причинам, что и при определении h_6 вместо количества переданной информации будет использоваться объем переданного сообщения. Поскольку этот объем может изменяться в процессе преобразования сообщений и, в частности, при кодировании, то в случаях, когда возможны сомнения, будет специально оговариваться относительно какого именно объема вычисляется γ .

В настоящей главе эта величина будет всегда исчисляться относительно объема сообщения на входе радиолинии. Необходимо также обратить внимание на неоднозначность понятия «полоса частот занята сигналом», фигурирующего в определении γ . Когда речь идет об одноканальной передаче, под полосой понимают ширину спектра сигнала по уровню — 3 дБ (в настоящей главе она принимается равной B). Когда же речь идет о многоканальной передаче с использованием ряда несущих, то в полосу включают еще и защитные интервалы. С таким случаем мы встретимся в следующей главе.

Оценка энергетической эффективности сигналов при m -ичной передаче осложняется отсутствием точных выражений для потенциальной помехоустойчивости не двоичных методов манипуляции. Ограничимся некоторыми простыми оценками, которые непосредственно следуют из данных гл. 3. Пусть используется m -ичная манипуляция. Выберем некоторую пару возможных ЭС $s_i(t)$ и $s_j(t)$ и рассмотрим двоичную манипуляцию, использующую только эти сигналы. В соответствии с (3.16) вероятность ошибки при идеальном когерентном приеме в канале с БГШ составит

$$P_{\text{ош}} = \Phi(d_{ij} / \sqrt{2N_0}), \quad (5.1)$$

где величина

$$d_{ij} = \sqrt{\int_0^{T_s} [s_i(t) - s_j(t)]^2 dt} \quad (5.2)$$

может рассматриваться как евклидово расстояние между сигналами с номерами i и j .

На рис. 5.1 условно точками представлено множество возможных ЭС и показано расстояние d_{ij} между соответствующей парой этих сигналов. На рисунке отмечена также пара сигналов, расстояние между которыми минимально из таковых для всех возможных пар. Это расстояние обозначено через d_0 .

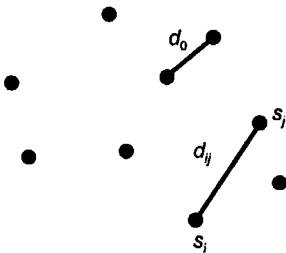


Рис. 5.1. Элементарные сигналы и расстояния между ними

Как следует из изложенной выше теории идеального приема, в каналах с БГШ наиболее часто данный переданный сигнал трансформируется в один из ближайших к нему возможных ЭС. Поэтому если число таких ближайших сигналов составляет в среднем M , то вероятность ошибочного приема ЭС можно ориентировочно оценить величиной

$$P_{\text{ошз}} = M\Phi(d_0 / \sqrt{2N_0}). \quad (5.3)$$

По поводу этой оценки, которой мы ниже будем постоянно пользоваться, сделаем несколько замечаний. Во-первых, о ее точности следует судить по точности определения уровня сигнала, требуемого для обеспечения заданной вероятности ошибки. Именно такой подход соответствует нуждам практики и только при таком подходе можно считать (5.3) приемлемой оценкой. Дело в том, что, поскольку функция ошибок очень быстро (экспоненциально) убывает с ростом аргумента, оказывается, что даже малое изменение d_0 может привести к изменению $P_{\text{ошз}}$ во много раз. Во-вторых, по тем же соображениям, влияние числа M на приведенную оценку оказывается второстепенным, особенно когда

требуемое значение вероятности ошибки достаточно мало. Для первоначальной оценки иногда вообще полагают $M = 1$, ограничиваясь лишь сопоставлением минимальных евклидовых расстояний между элементарными сигналами для сравниваемых вариантов. Наконец, последнее замечание призвано пояснить количественную зависимость между вероятностями ошибки в приеме ЭС ($P_{\text{ошз}}$) и в воспроизведении бит модулирующего цифрового сигнала ($P_{\text{ошб}}$). Эта связь зависит от того, как установлено соответствие между блоками (комбинациями) из k последовательных бит модулирующего сигнала и ЭС. Это соответствие именуется *манипуляционным кодом*.

При рациональном манипуляционном кодировании близким ЭС должны назначаться близкие блоки сообщения. При этом близость ЭС должна оцениваться по евклидовому расстоянию между ними, а блоков — по соответствующему хеммингову расстоянию. Для ФМ известны манипуляционные коды (коды Грея), при которых отличие в блоках, соответствующих сигналам с расстоянием d_0 составляет 1 бит. Тогда

$$P_{\text{ошб}} = P_{\text{ошз}}. \quad (5.4)$$

Если же использовать случайный манипуляционный код, то

$$P_{\text{ошб}} = 0,5 k P_{\text{ошз}}, \quad (5.5)$$

поскольку при ошибочном приеме посылки в среднем ошибочно воспроизводится половина символов соответствующего блока.

Учитывая, что во всех случаях речь может идти об отличии $P_{\text{ошб}}$ от $P_{\text{ошз}}$ всего в несколько раз, в рассматриваемом приближении можно всегда полагать, что выполняется (5.4), что мы ниже и примем.

В соответствии с вышеизложенным при использовании биортогонального ансамбля сигналов в полосе Δf при ЭС длительностью T_3 обеспечивается битовая скорость передачи

$$D = \frac{\log_2(2m)}{T_3} = \Delta f \frac{2 \log_2(2m)}{m} \quad \text{при } m \geq 2 \text{ и четном.} \quad (5.6)$$

Максимальная скорость достигается при $m = 2$ и составляет

$$D = 2\Delta f \quad (5.7)$$

и достигается при ФМ4.

Чтобы сформировать на базе сигналов ФМ4 метод передачи информации, который удобно использовать в качестве опорного для сравнения с другими вариантами, учтем еще возможности, даваемые помехоустойчивым кодированием (см. гл. 4). При этом будем исходить из блок-схемы тракта передачи, представленной на рис. 5.2.



Рис. 5.2. Структурная схема тракта передачи

Входящий в эту схему кодер представляет собой обратимый преобразователь сообщения, изменяющий его объем. Будем полагать, что и сообщение, и кодовые комбинации представлены в двоичном виде, но их тактовые частоты различны, в связи с чем и необходимо различать символы информации («информационные биты») и кодовых комбинаций («кодовые биты»).

Примем, например, что цифровое сообщение перед модулятором кодируется сверточным кодом с длиной кодового ограничения 7 и скоростью $R = 1/2$, эффективность которого была рассмотрена в 4.5. Тогда для опорного варианта получим:

– скорость передачи информации

$$C_0 = RD = 0,5D = \Delta f ; \quad (5.8)$$

– удельную скорость передачи информации

$$\gamma_0 = \frac{C_0}{\Delta f} = 1 ; \quad (5.9)$$

– величину нормированного отношения сигнал/шум, необходимого для обеспечения вероятности ошибки в приеме информационного бита $P_{\text{ошиб}} = 10^{-5}$:

$$h_{60}^2 = E_6 / N_0 = 2,95 \text{ раза} = 4,6 \text{ дБ}, \quad (5.10)$$

где E_6 — энергия сигнала, приходящаяся на информационный бит.

Перечисленные параметры являются базовыми для сравнения с другими вариантами выбора сигналов, к которым приходится прибегать, если при заданной полосе пропускания требуется обеспечить битовую скорость, большую предела, определяемого выражением (5.7). В этих случаях уже не может быть использован ансамбль ортогональных или биортогональных сигналов. Приходится применять сигналы с нормированным коэффициентом корреляции r (см. обозначения, относящиеся к (3.10)), лежащим между 0 и 1.

5.1.2. Построение полосно-эффективных сигналов

Ограничимся вопросами выбора таких сигналов для сигнальных пространств размерности $m = 2$, поскольку именно они наиболее часто реализуются на практике. Как отмечалось выше, такие сигналы могут быть представлены векторами на плоскости. Применим также представление таких сигналов точками плоскости, рассматриваемыми как концы соответствующих векторов.

Чаще всего используются два варианта ансамблей сигналов:

- ансамбль сигналов фазовой манипуляции (ФМ); геометрическое представление n -ичного ансамбля показано на рис. 5.3, а;
- ансамбль сигналов амплитудно-фазовой манипуляции (АФМ); геометрическое представление варианта 16-ричного ансамбля дано на рис. 5.3, б.

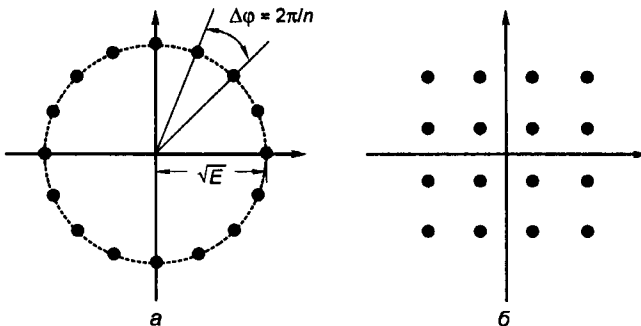


Рис. 5.3. Ансамбли полосно-эффективных сигналов:
а, б — геометрическое представление n -ичного и 16-ричного соответственно

Для ФМ ансамбль сигналов является нормированным (все сигналы имеют равные энергии E), и минимальное евклидово расстояние между ними как функция от n составляет

$$d_0(n) = 2\sqrt{E} \sin(\pi/n) = 2\sqrt{E_6 \log_2 n} \sin(\pi/n). \quad (5.11)$$

Из этого выражения следует, что расстояние d_0 быстро уменьшается с ростом n . Поэтому в качестве первого шага в направлении повышения удельной скорости передачи γ можно сохранить значение n , принятое для базового варианта, но отказаться от кодирования (принять $R = 1$). Это даст вариант с параметрами: $\gamma = 2$ и $d_0 = \sqrt{2E} = 2\sqrt{E_6}$. В соответствии с (5.3) для обеспечения вероятности ошибки 10^{-5} потребуется $d_0 / \sqrt{2N_0} = \sqrt{2} \sqrt{E_6 / N_0} = \sqrt{2} h_6 = 4,27$.

Иначе говоря, требуемое значение приведенного отношения сигнал/шум составит $h_6 = 9,6$ дБ. (5.12)

Таким образом, исключение кодирования приводит к возрастанию спектральной эффективности в два раза за счет снижения энергетической на 5 дБ.

Если теперь перейти к использованию ФМ с большими значениями n , но без кодирования, то из (5.3) и (5.11) найдем: (5.13)

$$\gamma(n) = \log_2 n = k; \quad (5.13)$$

$$h_6(n) = \frac{n}{\sqrt{\log_2 n}} = \frac{2^k}{\sqrt{k}}. \quad (5.14)$$

Отметим, что последнее соотношение является приближенным и может использоваться при $n \geq 8$.

Сопоставляя (5.14) со значением h_6 для базового варианта, определим величину энергетического проигрыша для $n \geq 8$:

$$\eta(n) = (h_6 / h_{60})^2 = 0,355 \frac{2^{2k}}{k}. \quad (5.15)$$

В табл. 5.1 представлены показатели энергетической и спектральной эффективности ФМ при различных n при отсутствии кодирования. При этом для $n \leq 4$ использованы точные данные, полученные в гл. 3, а для больших значений n данные получены с использованием оценок настоящего раздела.

Таблица 5.1. Энергетическая и спектральная эффективность многопозиционной ФМ

N	2	4	8	16	32	64
η , дБ	5	5	8,8	13,6	18,6	23,85
γ , бит/с/Гц	1	2	3	4	5	6

Перейдем к рассмотрению случая использования ансамблей элементарных сигналов АФМ. Такие сигналы уже не являются нормированными. Если полагать, что все сигналы ансамбля используются равновероятно, то для $n = 16$ ($k = 4$) можно подсчитать, что минимальное евклидово расстояние такого ансамбля составляет

$$d_0 = \sqrt{\frac{2E}{5}} = \sqrt{\frac{8E_6}{5}}, \quad (5.16)$$

где E — средняя мощность сигнала АФМ; $E_6 = E/k = 0,25E$ — средняя энергия, приходящаяся на передаваемый бит. Аналогично тому, как это было сделано выше, найдем, что в данном случае для обеспечения вероятности ошибки 10^{-5} необходимо $h_6 = 13,5$ дБ, что соответствует

$$\eta(16) = 9 \text{ дБ}. \quad (5.17)$$

Для $n = 16$ ансамбль АФМ энергетически существенно эффективнее ансамбля ФМ (на 4,6 дБ). Отметим, что для $n = 4$ эти ансамбли совпадают. Таким образом, с ростом n относительная эффективность АФМ по отношению к ФМ возрастает.

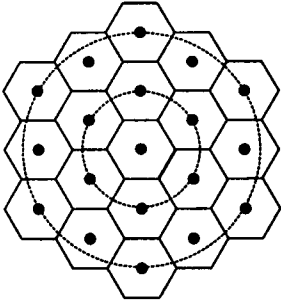


Рис. 5.4. Размещение сигналов по гексагональной решетке

Заданным минимальным расстоянием на круге. В пределе, когда отношение радиуса круга к заданному расстоянию достаточно велико, наиболее плотной оказывается гексагональная решетка (рис. 5.4).

Сигнальные точки расположены здесь в центрах правильных шестиугольников, которыми замощена плоскость. Если точку, расположенную в центре плоскости, выбрать центральной, то остальные шестиугольники образуют кольца вокруг данной точки. Первое такое кольцо содержит шесть шестиугольников, центры которых отстоят от нулевой точки на расстоянии d_0 . Если ограничиться только этими точками, то получим ансамбль из семи точек со средней энергией, близкой к $E = d_0^2$.

Можно расширить ансамбль, прибавив к нему точки, относящиеся еще к нескольким кольцам шестиугольников. Легко видеть, что каждое следующее кольцо имеет на шесть точек больше предыдущего, так что если используется m колец (нулевая точка кольцом не считается), то общее число точек ансамбля составит

$$n = 1 + \sum_{i=1}^m 6i = 3m(m+1) + 1. \quad (5.18)$$

Значения n в зависимости от m образуют последовательность 1, 7, 19, 37, 61 и т.д. Использовать ансамбли с таким количеством сигналов неудобно, поскольку они не составляют целую степень 2. Приходится не использовать некоторую часть сигнальных точек. Так, при $m = 3$ целесообразно реализовать только 16 точек, опуская три точки наружного кольца с максимальной энергией. При этом средняя энергия элементарной посылки составит

$$E = 4E_6 = \frac{39}{19} d_0^2 = 2,05 d_0^2. \quad (5.19)$$

Для $m = 4$ целесообразно использовать 32 сигнала, опустив пять сигналов максимальной энергии. При этом получим

$$E = 5E_6 = \frac{141}{32} d_0^2 = 4,41 d_0^2. \quad (5.20)$$

Так же как и выше, найдем, что, в соответствии с этими данными для гексагональной решетки с $n = 16$ необходимо обеспечить

$$h_6 = 12,61 \text{ дБ, т.е. } \eta = 7,01 \text{ дБ,} \quad (5.21)$$

а для $n = 32$

$$h_6 = 16,15 \text{ дБ, т.е. } \eta = 11,55 \text{ дБ.} \quad (5.22)$$

Причины данного явления становятся ясными при рассмотрении рис. 5.3. Действительно, круг радиуса \sqrt{E} на рассматриваемой плоскости отражает все множество сигналов с энергией меньшей E . Ясно, что ансамбль сигналов будет тем более эффективным, чем больше сигнальных точек с заданным минимальным расстоянием d_0 удастся разместить в этом круге. Очевидно, что при не слишком малых n вариант размещения, используемый при АФМ, в этом смысле, предпочтительнее ФМ. Однако те же рассуждения приводят к выводу, что оба эти варианта не являются оптимальными. Наиболее выгодным должен оказаться вариант соответствующий наиболее плотному размещению точек с

Сопоставляя эти данные с приведенными выше убеждаемся, что действительно гексагональный ансамбль сигналов АФМ более эффективен, чем прямоугольный (при $n = 16$ энергетический выигрыш составляет 1,99 дБ) и, тем более, чем ансамбль ФМ (при $n = 16$ выигрыш 6,59 дБ). Отметим, что для $n = 4$ ансамбли АФМ и ФМ совпадают, а гексагональный фактически не существует.

5.2. Общие данные о сигнальном кодировании

В предыдущем параграфе были определены характеристики эффективности, которые могут быть достигнуты при использовании различных ансамблей сигналов объемом больше четырех в двумерном сигнальном пространстве. Предполагалось, что с целью достижения максимальной спектральной эффективности, помехоустойчивое кодирование перед модулятором не используется. Оказалось, что возникающие выигрыши в спектральной эффективности оплачиваются резким снижением энергетических показателей. В связи с этим возникает мысль, оставаясь в рамках полосно-эффективных ансамблей сигналов, применить кодирование, с тем чтобы за счет некоторого снижения γ существенно уменьшить энергетические потери.

Для выбора соответствующих кодов и оценки их эффективности в данном случае нельзя непосредственно воспользоваться материалами гл. 4. Дело в том, что на полосно-эффективные ансамбли сигналов оценки выигрыша от помехоустойчивого кодирования, приведенные в гл. 4, не распространяются.

Поясним это обстоятельство подробнее. Подавляющая часть кодов, рассмотренных в гл. 4, преобразует двоичные информационные сообщения в двоичные кодовые комбинации, а приведенные там оценки их эффективности основаны на определении расстояний между парами используемых кодовых комбинаций (спектров расстояний кодов). Евклидово расстояние между двоичными комбинациями пропорционально квадратному корню из расстояния Хэмминга между ними, т.е. числа отличающихся символов в данных комбинациях. В тракте передачи (см. рис. 5.2) кодовые комбинации преобразуются в последовательности ЭС, которые мы, для краткости, будем именовать *кодowymi цепочками*. Интерес представляет энергетический выигрыш от кодирования, выраженный именно в средней мощности соответствующего радиосигнала, а не кодовой комбинации. Поэтому оценки эффективности кодирования, приведенные в гл. 4, справедливы только для тех методов манипуляции, при которых евклидово расстояние между любой парой кодовых комбинаций пропорционально этому расстоянию между соответствующими им кодовыми цепочками.

Такая пропорциональность реализуется при любых двоичных методах манипуляции. Что же касается недвоичной манипуляции, то указанная пропорциональность обычно не имеет места. Однако она сохраняется в случае ФМ4 с манипуляционным кодом Грея. Действительно, здесь каждому блоку из двух кодовых бит ставится в соответствие своя элементарная посылка с энергией E . При этом блокам, отличающимся в одном бите, соответствуют ортогональные посылки (евклидово расстояние $d_1 = \sqrt{2E}$), а отличающимся в двух битах — противоположные посылки ($d_2 = 2\sqrt{E}$). Таким образом, здесь евклидово расстояние между сигналами действительно пропорционально таковому между кодовыми комбинациями. Легко убедиться, что в других рассмотренных выше вариантах недвоичной манипуляции указанное свойство не выполняется, что приводит к необходимости специального выбора кодов для каждого метода манипуляции и оценки их эффективности.

Задачей оптимизации формирования ансамбля кодовых цепочек должно быть или достижение как можно большего минимального евклидова расстояния между цепочками при

заданном их числе в ансамбле либо, наоборот, максимизация этого числа при заданном минимальном расстоянии. Во всяком случае, соответствующие процедуры должны учитывать способ формирования радиосигналов, присущий используемому в радиолинии методу манипуляции. По этой причине рассматриваемые методы кодирования именуется сигнальными, а возникающие в результате их реализации радиосигналы — сигнально-кодowymi конструкциями (СКК).

При сигнальном кодировании структура передающего тракта радиолинии (см. рис. 5.2) изменяется за счет объединения кодера и модулятора в единое устройство, которое мы будем именовать *сигнальным кодером*, сохраняя привычное наименование кодер за соответствующим преобразователем двоичных символов сообщения. Таким образом, сигнальный кодер ставит в соответствие последовательностям информационных бит, поступающих на его вход, цепочки элементарных сигналов, создаваемых им на выходе. При этом специально учитывается спектр расстояний между используемыми элементарными сигналами, так чтобы обеспечить максимальное евклидово расстояние между ближайшими используемыми цепочками.

Для решения поставленной задачи необходимо, прежде всего, так или иначе упорядочить элементарные сигналы по их взаимным расстояниям. Это достигается поэтапным разбиением всего ансамбля элементарных сигналов на вложенные множества сигналов, при котором весь ансамбль делится вначале на несколько множеств сигналов, каждое из которых делится на несколько подмножеств и т.д. Такое разбиение производится исходя из расстояний между множествами элементарных сигналов, причем под *расстоянием между двумя множествами сигналов* понимается минимальное расстояние между парой сигналов, один из которых выбран из первого множества, а другой — из второго. Разбиение ансамбля на вложенные множества требует, чтобы расстояния между любой парой множеств, образующихся на j -м шаге разбиения ρ_j , было бы одинаковым, причем на каждом следующем шаге данное расстояние по крайней мере не убывало бы, т.е. $\rho_j \leq \rho_{j+1}$. Обозначают такое разбиение ансамбля сигналов на вложенные множества выражением

$$X = (m; \rho_1, \rho_2, \dots, \rho_S; K_1, K_2, \dots, K_S), \quad (5.23)$$

где m — как и выше, размерность пространства сигналов, разделение которых произведено в $S - 1$ шагов; K_j — число подмножеств получающееся на j -м шаге разбиения из каждого множества предыдущего уровня; ρ_j — евклидово расстояние между различными множествами j -го уровня; K_S — число элементарных сигналов в каждом из финальных (образовавшихся после последнего разделения) подмножеств; ρ_S — минимальное евклидово расстояние между парой сигналов, входящих в одно и то же такое подмножество

Примеры

Рассмотрим множество биортогональных сигналов. В соответствии с (5.23) его можно записать, не используя разбиений ($s = 1$), как $X = (m; \sqrt{2E}; 2m)$. Возможно его одноуровневое ($s = 2$) разбиение вида $X = (m; \sqrt{2E}, 2\sqrt{E}; m, 2)$, т.е. на m пар противоположных сигналов. Можно использовать и другое разбиение на два множества ортогональных сигналов, причем сигналы второго множества противоположны сигналам первого: $X = (m; \sqrt{2E}, \sqrt{2E}; 2, m)$. Здесь и ниже под E понимается средняя энергия ЭС, причем принимается, что все такие сигналы передаются равновероятно.

Разложение ансамбля элементарных сигналов ФМ8 на вложенные множества

$$X = (2; \sqrt{0,586E}, \sqrt{2E}, 2\sqrt{E}; 2, 4, 2) \quad (5.24)$$

иллюстрирует рис. 5.5. Пути разбиения обозначены цифрами 0 или 1. В соответствии с нумерацией путей, ведущих к данному конкретному сигналу, можно сформировать и его

номер. В данном случае он будет представлять собой вектор с тремя двоичными координатами. Условимся записывать координаты, считываемые на рис. 5.5 сверху вниз (от общего к частному), в строку справа налево. Так, сигнал с номером 0 в третьем по порядку подмножестве нижнего уровня имеет векторный номер 0,0,1. Если рассматривать этот набор чисел как трехразрядное двоичное число (не обращая внимания на запятые), то в восьмеричном виде его можно записать одной цифрой 1. Полученную таким образом цифру будем считать порядковым номером сигнала и, чтобы отличить ее от координат векторного номера этого сигнала, выделим жирным шрифтом> 1. Полученную таким образом цифру будем считать порядковым номером сигнала и, чтобы отличить ее от координат векторного номера этого сигнала, выделим жирным шрифтом

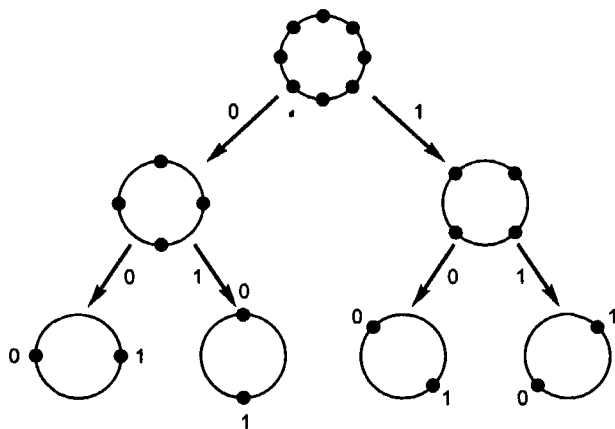


Рис. 5.5. Разложение ансамбля сигналов ФМ8

На рис. 5.6 представлено одно из возможных разбиение на вложенные множества ансамбля сигналов КАМ16. Термин квадратурная амплитудная манипуляция (КАМ), часто используется как аналог АФМ.

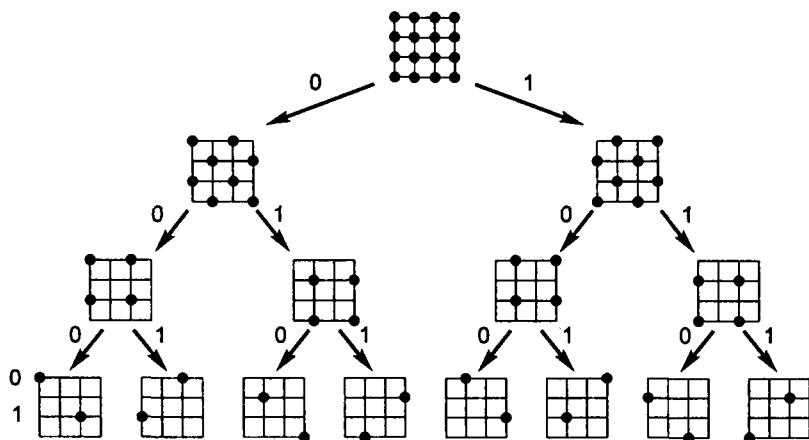


Рис. 5.6. Разложение ансамбля сигналов КАМ16

Формула такого разбиения имеет вид

$$X = (2; \sqrt{0,4E}, \sqrt{0,8E}, \sqrt{1,6E}, 2\sqrt{0,8E}; 2, 2, 2, 2). \quad (5.25)$$

Векторные номера сигналов представляют здесь четырехразрядное двоичное число, а порядковые номера — шестнадцатеричные числа. В соответствии с числом разрядов векторного номера это разбиение именуют четырехуровневым.

Можно строить и другие разбиения ансамбля КАМ16, среди которых отметим двухуровневое разбиение с формулой

$$X = (2; \sqrt{0, 4E}, \sqrt{1, 6E}; 4, 4). \quad (5.25a)$$

Отличие этого разбиения от предыдущего заключается в том, что ансамбль КАМ16 сразу делится на четыре четырехсигнальных множества, аналогичных тем, которые получались выше на второй ступени деления. Далее каждое образовавшееся подмножество рассматривается как четыре отдельных элементарных сигнала. При таком делении векторный адрес сигнала задается двухкоординатным вектором, значения координат которого выражаются четверичными числами.

Приведем еще разбиение на вложенные множества ансамбля сигналов, построенного на базе гексагональной решетки (см. рис. 5.4). При этом используем 18 элементарных сигналов, опустив центральный (нулевой) сигнал. Соответствующий ансамбль сигналов представлен на рис. 5.7, причем около каждой сигнальной точки приведен ее векторный адрес, что полностью описывает предлагаемое разбиение. Формула этого разбиения такова:

$$X = \left(2; \sqrt{\frac{3E}{8}}, 3\sqrt{\frac{E}{8}}, 3\sqrt{\frac{3E}{8}}; 3, 3, 2 \right). \quad (5.26)$$

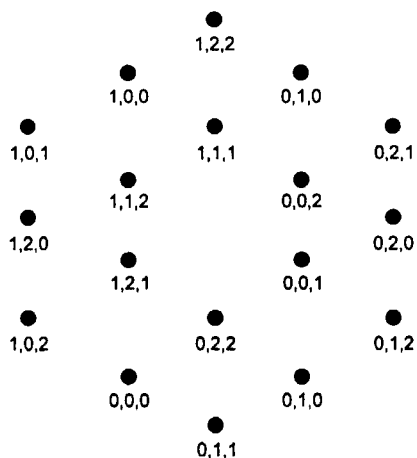


Рис. 5.7. Разбиение ансамбля сигналов, основанного на гексагональной решетке

Формула разбиения ансамбля сигналов на вложенные множества содержит достаточно полные данные о выбранных ЭС и предназначена для использования кодером в процессе выбора множества максимально разнесенных по евклидовой расстоянию кодовых комбинаций с учетом спектра расстояний между элементарными сигналами. К сожалению, не известны регулярные пути выбора алгоритмов работы таких кодеров, хотя в литературе и приводится много вариантов таких алгоритмов, полученных эвристическим путем с широким использованием методов перебора и машинного моделирования. Ограничимся описанием некоторых из этих вариантов.

5.3. СКК на базе решетчатых сигнальных кодов

Наибольшее число предложенных в литературе способов сигнального кодирования основано на использовании решетчатого кодирования. Решетчатый сигнальный кодер (РСК), так же как и сверточные кодеры, описанные в гл. 4, с точки зрения математической логики представляют собой *машину Тьюринга*. Такими машинами называют устройства с конечным числом состояний, реакция которых на то или иное входное воздействие сводится к по-

явлению на выходе соответствующего выходного сигнала, а также переходу устройства в новое состояние. Как выходной сигнал, так и новое состояние зависят от входного сигнала и состояния, в котором устройство находится в момент воздействия. Соответственно сигнальный кодер описывается направленным графом, где точками отмечены возможные состояния, стрелками показаны переходы между ними. Каждая такая стрелка соответствует определенному входному воздействию и определенному выходному сигналу. Вместо направленного графа, как и при сверточном кодировании, часто используется решетчатая диаграмма, представляющая собой временную развертку направленного графа и показывающая порядок смены состояний кодера во времени при поступлении тех или иных сообщений.

Работа такого сигнального кодера при поступлении на вход информационной последовательности бит, в интересующем нас случае, может быть описана следующим образом: входная последовательность разбивается на блоки по k_n бит и после каждого такого блока происходит смена состояния кодера, сопровождающаяся появлением на его выходе некоторого ЭС. Число возможных ЭС $n = 2^k$, причем $k > k_n$. Отношение $R = k_n/k$ является скоростью кода и определяет избыточность, связанную с кодированием.

Число состояний сигнального кодера обозначим через g . Из каждого состояния возможно $n_n = 2^{k_n}$ переходов в соответствии с числом различных блоков входного воздействия. При этом, если $g < n_n$, то должно быть предусмотрено по несколько переходов между некоторыми парами состояний (параллельные переходы); эти переходы должны отличаться входными блоками и выходными элементарными сигналами. Параллельные переходы могут существовать и при $g \geq n_n$. Отметим еще, что полной характеристикой того или иного состояния является перечень переходов с указанием состояния, в которое этот переход направлен, и соответствующими ему входным блоком и выходным сигналом. Условимся описывать входные блоки десятичными числами соответствующими комбинации двоичных чисел, если записывать их справа налево в порядке появления во времени. Так, запись блока 110 означает, что вначале появился символ 0, потом 1 и последним снова 1. Десятичное обозначение блока будет 6. Напомним, что выше предложен и вариант нумерации элементарных сигналов при многопозиционных ФМ и КАМ десятичными числами, но напечатанными жирным шрифтом.

Выше уже отмечалось, что выбор подходящей для данного ансамбля элементарных сигналов решетчатой диаграммы (алгоритма кодирования) осуществляется в значительной степени путем перебора. Тем не менее установлены некоторые эвристические правила, которыми следует руководствоваться при таком переборе. Эти правила значительно сокращают процедуру перебора. Сформулируем эти правила применительно к ансамблю сигналов ФМ8 (см. рис. 5.5). Они сводятся к следующим положениям:

а) если в решетчатой диаграмме используются параллельные переходы, то соответствующие им сигналы должны принадлежать одному и тому же финальному множеству, т.е. иметь векторные номера, отличающиеся только в первой координате;

б) все переходы из некоторого состояния должны соответствовать сигналам, принадлежащим одному из множеств первого уровня разделения, т.е. должны иметь векторные номера, с одинаковой последней координатой;

в) все ЭС должны использоваться одинаково часто.

Остановимся еще на способах оценки энергетической эффективности СКК. В соответствии с (3.16) она определяется минимальным евклидовым расстоянием между цепочками элементарных сигналов, возникающими на выходе модулятора при передаче различных информационных последовательностей. Как правило, рассматривают две информационные последовательности, совпадавшие до начального момента времени $t = 0$, в которых

в данный момент появляются различные символы. Соответствующие этим последовательностям цепочки элементарных сигналов тоже совпадают до $t=0$ и начинают расходиться после этого момента. Однако, если на последующих тактах информационные символы обеих последовательностей снова начнут совпадать, то выходные цепочки опять сольются. Таким образом, понятие евклидова расстояния между цепочками приобретет очевидный смысл. Если теперь при заданном ансамбле сигналов и методах кодирования варьировать различие между двумя рассматриваемыми информационными последовательностями, то можно установить, каково минимальное евклидово расстояние между двумя цепочкам, отличающимися на некотором конечном временном интервале. Эта величина именуется свободным расстоянием кода r_c .

При безыбыточной передаче r_c совпадает с минимальным евклидовым расстоянием между сигналами используемого ансамбля d_0 . Кодирование, вообще говоря, увеличивает это расстояние, чем и определяется его полезный эффект. Он должен определяться выражением (5.1) при подстановке в него величины r_c вместо d_0 .

Рассмотрим теперь некоторый конкретный пример РСК. Пусть при заданной скорости передачи информации полоса выделенных рабочих частот такова, что может быть использована манипуляция ФМ4, но без избыточного помехоустойчивого кодирования. Как следует из изложенного в предыдущем параграфе, такая передача соответствует снижению помехоустойчивости (по сравнению с базовым вариантом) на 5 дБ (см. (5.12)). Перейдем к использованию ФМ8. Увеличение числа возможных элементарных сигналов может быть реализовано для:

- увеличения в 1,5 раза скорости передачи информации, а вместе с ней и γ , без расширения полосы, но со снижением помехоустойчивости уже на 8,8 дБ (см. табл. 5.1);
- введения помехоустойчивого кодирования без изменения скорости передачи информации с целью повышения помехоустойчивости.

Проанализируем последнюю возможность. При этом будем использовать простейшие варианты РСК.

Используем решетчатую диаграмму с четырьмя состояниями, представленную на рис. 5.8.

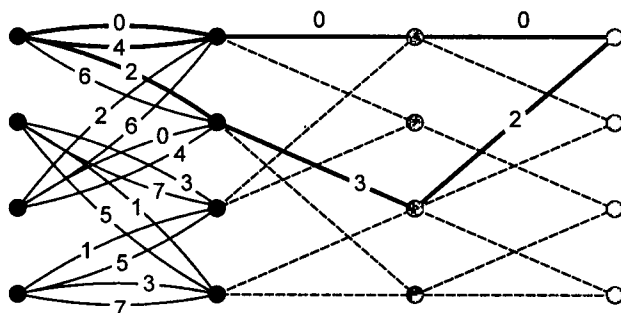


Рис. 5.8. Решетчатая диаграмма сигнального кодера

На диаграмме цифрами показаны порядковые номера сигналов, появляющихся на выходе модулятора при соответствующем переходе из предыдущего состояния в последующее. Структура решетчатой диаграммы соответствует приведенным выше эвристическим рекомендациям. Действительно, все сигналы переходов, исходящих из одного (любого) состоя-

ния принадлежат одному (из двух) множеству верхнего уровня (рекомендация «б»), поскольку двоичные записи порядковых номеров этих сигналов имеют одинаковые последние цифры. То же относится и к номерам сигналов, соответствующих входящим в данное состояние переходам.

Далее, в соответствии с рекомендацией «а», параллельным переходам всюду соответствуют сигналы одного финального множества (с одинаковыми двумя последними цифрами векторного номера). Наконец, каждый сигнал используется ровно в $1/8$ всех переходов решетки (рекомендация «в»). Таким образом, все рекомендации в данном случае выполняются. Следует заметить, что именно эти рекомендации приводят к появлению параллельных путей на решетчатой диаграмме. Действительно, исключение таких путей привело бы к необходимости создания переходов из любого состояния ко всем состояниям, что сделало бы невозможной выполнение рекомендации «б». Легко обнаружить, что эти рекомендации далеко не определяют однозначно все необходимые параметры СКК. Недостающие параметры определены в литературе путем перебора.

Для оценки энергетической эффективности СКК, описываемой решеткой на рис. 5.8, необходимо определить соответствующее свободное расстояние. Оно определяется с помощью решетки перебором различных возможных путей и анализом попарных евклидовых расстояний между ними. Поскольку код, описываемый решеткой, является линейным, нет необходимости перебирать все пары возможных цепочек (путей); в качестве базового можно выбрать один путь и определять расстояния всех остальных путей до базового. В качестве базового пути выберем, например, цепочку состоящую из последовательности сигналов 0 (верхняя горизонтальная линия на рисунке). Путем перебора можно убедиться, что, если не учитывать параллельных переходов, то наиболее близким к базовому пути является цепочка, образованная сигналами 2, 3, 2. В соответствии с (5.24) свободное расстояние составило бы при этом $\sqrt{4,586E}$. Однако, если учесть и параллельные переходы, то окажется, что есть и более близкие к базовой цепочки. Такой является цепочка, образованная сигналами 4, 0, 0. Она и определяет свободное расстояние рассматриваемой СКК, которое оказывается $\rho_c = 2\sqrt{E}$. Поскольку в данном случае каждый элементарный сигнал переносит 2 бита информации, $E = 2E_6$, так что

$$\rho_c = 2\sqrt{2E_6}. \quad (5.27)$$

В результате, используя (5.1), находим оценку

$$P_{\text{ом}} = \Phi(\rho_c / \sqrt{2N_0}) = \Phi(2h_6). \quad (5.28)$$

Сопоставляя этот результат с данными по помехоустойчивости ФМ4 без кодирования, можно заключить, что применение рассмотренной СКК, использующей ФМ8, позволяет без расширения полосы частот занимаемой сигналом повысить помехоустойчивость на 3 дБ. Тем не менее по отношению к базовому варианту (ФМ4 с кодированием) и здесь сохраняется известный энергетический проигрыш, составляющий 2 дБ. Зато удельная скорость передачи информации при рассмотренной СКК $\gamma = 2$ бит/с/Гц в два раза превосходит получаемую в базовом варианте. В этом проявляется достигаемый с помощью СКК обмен энергетической эффективностью на полосную.

Следует отметить, что все выводы получены безотносительно к конкретному информационному содержанию, приписываемому каждому переходу решетки. Отсюда можно заключить, что асимптотические показатели помехоустойчивости СКК не зависят от способов манипуляционного кодирования, т.е. способа назначения элементарным сигналам комбинаций информационных символов. Это не исключает влияния манипуляционного кодирования на помехоустойчивость при конечных, особенно при достаточно больших, вероятностях ошибки.

Приведем еще блок-схему кодера (рис. 5.9), реализующего решетку на рис. 5.8. Здесь представлен кодер, состоящий из регистра сдвига на три элемента памяти и двух сумматоров по mod 2. На вход кодера поступают параллельно два очередных информационных бита, при длительности их воздействия $2T$, где T — средняя длительность передачи информационного бита. Кодировается сверточным кодом только первый бит, в результате чего возникают два кодовых бита (на выходах кодера a и b). Второй информационный бит не кодируется и непосредственно поступает на контакт $в$ как третий кодовый бит. Таким образом, на два информационных бита кодер выдает три кодовых бита, так что скорость кода $R = 2/3$.

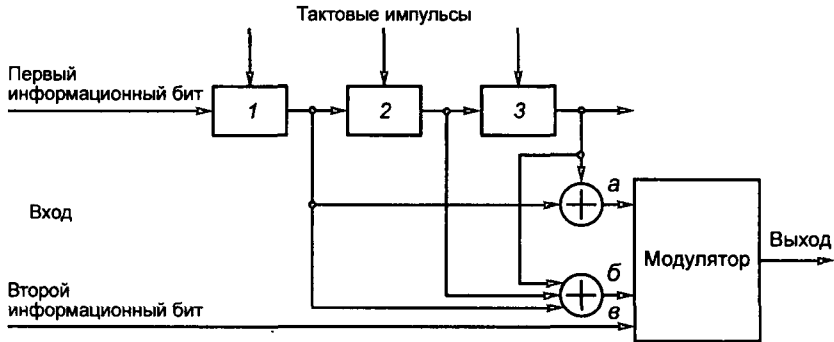


Рис. 5.9. Блок-схема сигнального кодера

Что касается модулятора, то на его выходе при поступлении каждой тройки кодовых бит выдается один из элементарных сигналов ФМ8. Манипуляционный код (зависимость порядкового номера выходного сигнала от набора поступивших кодовых бит) представлен в табл. 5.2.

Таблица 5.2. Манипуляционный код модулятора

Кодовая комбинация			Сигнал	Кодовая комбинация			Сигнал
a	b	$в$		a	b	$в$	
0	0	0	0	1	0	0	3
0	0	1	4	1	0	1	7
1	1	0	2	0	1	0	1
1	1	1	6	0	1	1	5

Легко убедиться, что СКК, формируемая устройством на рис. 5.9, при таком манипуляционном коде полностью соответствует решетке на рис. 5.8.

Как было показано выше, наличие в этой решетке параллельных путей ограничивает величину ее свободного расстояния. Поэтому представляется логичным пойти на известное усложнение кодера рассматриваемой СКК, перейдя к использованию решетки с восьмью состояниями. Соответствующая СКК представлена решеткой на рис. 5.10, a . Для упрощения слева от точки каждого состояния выписано некоторое множество сигналов ФМ8, возникающих при переходах, исходящих из этого состояния. Порядок следования сигналов (слева направо) соответствует порядку следования на диаграмме исходящих переходов (сверху вниз). На решетке показаны также жирными линиями два пути, евклидово расстоя-

ние между которыми минимально. Приведены и порядковые номера элементарных сигналов, образующих эти пути. Указанные данные позволяют установить, что в данном случае

$$\rho_c = 2,141\sqrt{E} = 2,141\sqrt{E_c}. \quad (5.29)$$

Сопоставляя это соотношение с (5.27), устанавливаем, что использование кодера с восьмью состояниями вместо четырех повышает асимптотическую помехоустойчивость СКК с ФМ8 на 0,6 дБ.

Остановимся еще на возможности использования РСК для построения СКК для КАМ16. Выбранный описанными выше методами вариант соответствующей СКК показан на рис. 5.10, б.

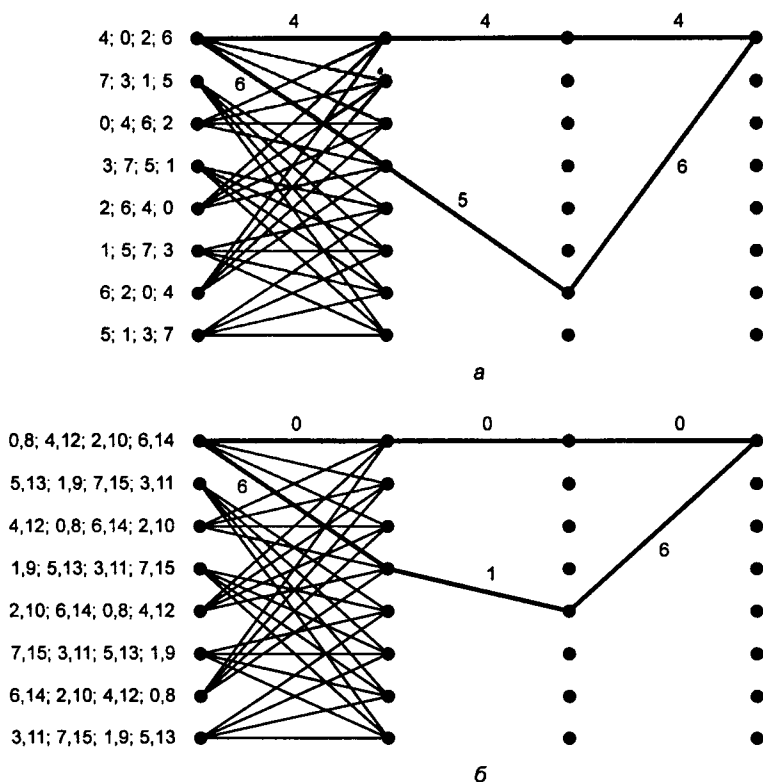


Рис. 5.10. Вариант СКК с ФМ8 (а) и с КАМ16 (б)

При построении рассматриваемой СКК принято, что каждый переход соответствует передаче трех информационных бит; поскольку КАМ16 имеет 16 сигналов, то это означает, что при кодировании на каждые три информационных бита добавляется один кодовый, так что скорость кода $R = 3/4$, а $\gamma = 3$ бит/с/Гц, т.е. в 1,5 раза больше, чем в рассмотренном выше варианте с ФМ8. Поскольку из каждого состояния должно исходить восемь путей (число различных комбинаций из трех информационных символов), то все показанные на рис. 5.10, б переходы в действительности являются параллельными; на рисунке две параллельные линии переходов заменены одной, чтобы не затемнять изображение. Порядковые номера элементарных сигналов, появляющиеся на выходе при каждом переходе, перечисле-

ны в левом столбце цифр, по аналогии с тем, как это было принято на рис. 5.9. Поскольку здесь каждый переход является параллельным, он соответствует паре сигналов. Номера сигналов, образующих такие пары, разделяются запятой, в отличие от различных пар, которые разделены точкой с запятой.

На рис. 5.10, б показаны также два пути, расстояние между которыми минимально. Это позволяет с использованием формулы разбиения ансамбля КАМ16 (5.25) определить

$$\rho_c = \sqrt{2E} = \sqrt{6E_6}. \quad (5.30)$$

Отметим, что евклидово расстояние между параллельными путям в данной СКК составляет $\rho = \sqrt{8E}$, так что оно не определяет свободное расстояние.

Сопоставляя полученные выше данные, устанавливаем, что переход от ФМ8 к КАМ16 при сохранении РСК с восьмью состояниями приводит к снижению помехоустойчивости на 1,25 дБ. Зато повышается эффективность использования полосы: величина γ возрастает в 1,5 раза. По сравнению с базовым вариантом γ больше в три раза за счет снижения помехоустойчивости на 3,25 дБ.

Можно также попытаться оценить эффективность СКК при неизменной полосе рабочих частот и скорости передачи информации, как метода повышения помехоустойчивости. Для этого сопоставим два варианта: передачу информации методом ФМ8 без кодирования и построенную выше СКК с КАМ16. В обоих случаях полная эффективность одинакова ($\gamma = 3$ бит/с/Гц), но помехоустойчивость второго варианта на 5,55 дБ выше.

5.4. СКК на базе обобщенных каскадных кодов

Каскадные коды, описанные в гл. 4, использовали два кода: один внешний и один внутренний. В отличие от этого обобщенные каскадные коды (ОКК) используют по несколько ступеней такого кодирования. Коды могут быть эффективно использованы для построения СКК. Ниже в общих чертах опишем соответствующие процедуры.

Для использования в рассматриваемой радиолинии задан ансамбль ЭС с формулой разбиения на вложенные множества вида (5.23). Примем, что подлежащая передаче информация представляет собой последовательность Q -ичных чисел, а количество подмножеств сигналов на каждом уровне разбиения являются целочисленными степенями Q , т.е. $a_j = \log_Q K_j$ — целые числа ($j = 1, 2, \dots, S$).

Пусть теперь для использования в качестве внешних задано S линейных блочных Q -ичных кодов вида $(m_0, k_j, d_j)_Q$ с одной и той же длиной кодовых комбинаций m_0 , числом информационных символов k_j и кодовым расстоянием d_j (как и прежде, $j = 1, 2, \dots, S$).

ОКК будем представлять в виде матрицы (см. рис. 5.11) с числом столбцов m_0 и числом строк $m_l = \sum_{j=0}^S a_j$, причем здесь принять $a_0 = 1$. Число S называется порядком ОКК.

Разобьем эту матрицу горизонтальными линиями на $S + 1$ подматрицу, так чтобы число строк в j -й подматрице составляло a_j (см. рис. 5.11, а). Запишем теперь в строки этой матрицы текущие информационные символы последовательно один за другим, но так чтобы они занимали только левые части строк, отделенные на рисунке вертикальными прямыми в подматрицах. После заполнения левой части очередной строки информационными символами, будем заполнять соответствующими проверочными символами ее правую часть, в соответствии с внешним кодом, предназначенным для данной подматрицы (номер кода совпадает с

номером подматрицы). Будем так продолжать до заполнения матрицы (кроме нулевой строки).

В заполненной матрице содержится $M = \sum_{j=1}^s a_j k_j$ информационных Q -ичных символов.

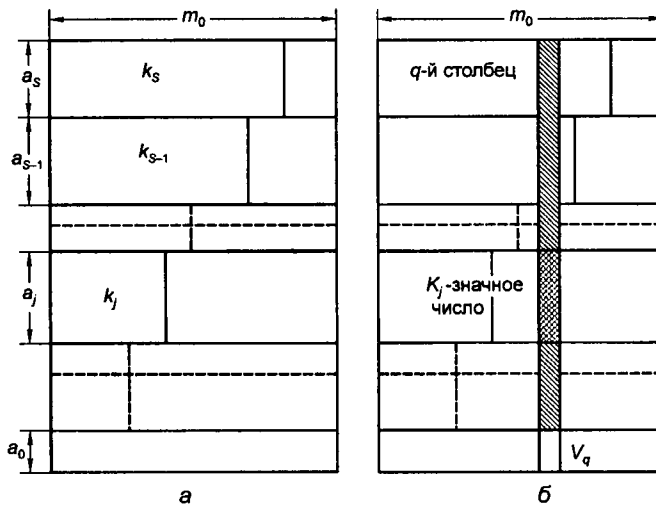


Рис. 5.11. Матрицы ОКК

В q -м столбце заполненной матрицы (см. рис. 5.11) в j -й подматрице содержится колонка Q -ичных чисел, включающая a_j чисел. Она может рассматриваться как одно K_j -значное число ($K_j = Q^{a_j}$). Это обстоятельство дает возможность передать весь q -й столбец матрицы одним элементарным сигналом из используемого ансамбля. В качестве такого сигнала V_q выберем тот, чей векторный номер совпадает с содержимым данного столбца матрицы. Этот сигнал и запишем в последнюю строку матрицы. Передав последовательно сигналы V_1, V_2, \dots, V_{m_0} , тем самым, передадим все содержимое матрицы, после чего вновь заполним ее и передадим следующую цепочку информационных символов и т.д.

Закончим на этом общее описание СКК данного типа и перейдем к оценке ее параметров.

Что касается удельной скорости передачи, то она составляет в данном случае

$$\gamma = \frac{M \log_2 Q}{m_0} \text{ бит/с/Гц.} \quad (5.31)$$

Остановимся еще на оценке свободного расстояния для данной СКК. Для его оценки необходимо сравнивать пары матриц описанного типа, соответствующие двум отличающимся хотя бы в одном символе информационным последовательностям. Для каждой такой пары целесообразно определять евклидово расстояние между последовательностями сигналов, занимающих последние строки матриц и искать среди этих расстояний минимальное. При сравнении матриц можно действовать следующим образом: будем последовательно сравнивать строки обеих матриц, начиная с верхней. Пусть впервые обнаружено отличие между некоторыми строками, принадлежащими j -м подматрицам. Тогда расстояние Хэмминга между этими строками не может быть меньше d_j , так что между последовательностями элементарных сигналов, занимающих последние строки этих матриц, расстояние Хэмминга также не менее d_j , причем евклидово расстояние между отличающимися сигналами обеих матриц, занимающих одну и ту же позицию, не менее ρ_j . Действительно, если раз-

личия впервые появляются в некоторой строке j -х подматриц, то это значит, что, начиная с уровня этой строки, появляются различия в d_j столбцах матриц, т.е. в векторных номерах соответствующих элементарных сигналов, причем, поскольку отличаются j -е координаты этих номеров, то расстояние между соответствующими сигналами не менее ρ_j . Последнее означает, что евклидово расстояние между последовательностями ЭС таких матриц не менее $\sqrt{d_j} \rho_j$. Из приведенных рассуждений следует:

$$\rho_c = \min_j \sqrt{d_j} \rho_j. \quad (5.32)$$

Приведенные соотношения будут использованы ниже при рассмотрении конкретных примеров построения СКК на базе ОКК.

Начнем со случая использования ансамбля КАМ16 с формулой разбиения (5.25а). В этом случае можно принять $Q = 4$; $S = 2$ и $a_j = 1$. Используем следующие внешние блочные РС-коды: $(4, 1, 4)_4$ для $j = 1$ и $(4, 4, 1)_4$ для $j = 2$. Последний код безызбыточный и означает просто, что кодирование в верхней подматрице вообще не используется. В соответствии с формулой разбиения и (5.25а) найдем для этого варианта $\rho_c = \sqrt{1,6E}$. Далее, определим, что заполненная матрица содержит 10 информационных бит. Таким образом, последовательность из четырех элементарных сигналов, имеющая суммарную энергию $4E$, переносит 10 бит информации, откуда следует $E = 2,5E_6$. Свободное расстояние теперь можно записать в виде $\rho_c = 2\sqrt{E_6}$. Используя параметры базового варианта (6.9) и (6.10), найдем, что полученная таким образом СКК имеет полосную эффективность в 2,5 раза больше, чем в базовом варианте, а помехоустойчивость на 5 дБ меньше. Если же сравнивать со случаем передачи КАМ16 без кодирования, то окажется, что рассматриваемая СКК имеет γ в 1,6 раза меньше, но помехоустойчивость больше на 4 дБ. Таким образом, ее использование действительно позволяет при сохранении полосы рабочих частот, занимаемых сигналом, существенно повысить помехоустойчивость за счет некоторого снижения полосной эффективности.

Другой пример СКК можно получить, если для того же ансамбля элементарных сигналов КАМ16 использовать другое разложение на вложенные множества, имеющее большее число уровней. Речь идет о формуле разбиения (5.25). Это разбиение отличается от предыдущего более детальной сортировкой элементарных сигналов по расстоянию. Интересно оценить в какой степени это позволяет улучшить СКК.

Подбор подходящих внешних кодов для данного случая позволил рекомендовать следующие коды (в порядке возрастания номера подматрицы j : $(4, 4, 1)_2$; $(4, 4, 1)_2$; $(4, 3, 2)_2$ и $(4, 1, 4)_2$. (Две верхние подматрицы внешними кодами не кодируются.) Для этого случая, как и выше, находим: $\gamma = 3$ бит/с/Гц и $\rho_c = \sqrt{4,8E_6}$, так что более детальный учет взаимных расстояний между сигналами КАМ16, позволяет на 20% повысить удельную скорость передачи информации γ при одновременном повышении помехоустойчивости на 0,8 дБ.

В заключение отметим, что приведенные в нем данные позволяют наглядно пояснить основные идеи, используемые при построении СКК.

В цифровой связи правильный прием элементарного сигнала полностью эквивалентен правильному определению на приемной стороне его номера в заранее известном наборе всех тех сигналов, которые могли быть переданы. Именно номер принятого сигнала определяет его конкретное информационное содержание. Первым шагом на пути построения СКК является формирование векторных номеров сигналов на основе разбиения всего ансамбля сигналов на вложенные множества. Сопоставляя построенные таким образом векторные

номера двух сигналов, можно оценить евклидово расстояние между ними: именно, если сопоставлять координаты соответствующих векторных номеров, начиная с самой правой, то расстояние между сигналами окажется тем большим, чем старше (левее) окажется тот разряд (координата), в котором впервые обнаружится различие векторных номеров этих сигналов.

Если на приеме используется идеальный демодулятор (приемник по максимуму правдоподобия), то ошибки в приеме будут сводиться, чаще всего, к регистрации вместо действительно переданного сигнала одного из ближайших к нему. Это означает, что ошибки в координатах принимаемых векторных номеров будут происходить не равновероятно: чаще они будут возникать в младших разрядах и сравнительно редко в старших. Этот факт позволяет реализовать следующий прием: не устанавливать однозначного соответствия между каждой координатой векторных адресов и информационными символами, а кодировать информационные сообщения в координаты элементарных сигналов, причем для каждого разряда это кодирование осуществлять отдельно, используя для младших разрядов более эффективные коды, чем для старших. Описанный выше алгоритм построения матриц и представляет собой вариант такого кодирования.

При удачном выборе набора внешних кодов в системе должна обеспечиваться приблизительно одинаковая вероятность ошибки в трансляции информационного символа, независимо от того на каком месте в матрице СКК он находится. При этом эффективно используются все данные о расстояниях между парами применяемых ЭС, а не только сведения о минимальном таком расстоянии, как это имеет место при обычном («цифровом») кодировании. В данном случае внешнее цифровое кодирование увязано со свойствами конкретного используемого ансамбля сигналов и корректирует его особенности и недостатки для достижения максимально возможной помехоустойчивости при заданной спектральной эффективности, или, наоборот, максимальной спектральной эффективности при заданной помехоустойчивости.

Глава 6

МНОГОСТАНЦИОННЫЙ ДОСТУП И МНОГОКАНАЛЬНАЯ СВЯЗЬ

6.1. Общая теория и классификация методов линейного разделения сигналов

6.1.1. Постановка задачи

Одной из ключевых задач техники связи признается обеспечение возможности использования имеющегося ресурса полосы рабочих частот в интересах как можно большего числа пользователей. При этом указанный ресурс должен служить для создания не одного канала связи между некоторой парой объектов, а ряда каналов между многими такими парами. Так, в полосах частот, выделенных для сотовой связи, в каждом районе, где такая связь организуется, общая среда распространения радиоволн используется в интересах создания целого ряда одновременно действующих каналов телефонной дуплексной связи. В спутниковой связи в качестве общего ресурса для организации каналов связи выступает ствол ретрансляции. В любой системе связи всегда имеется тот или иной общий для целого ряда каналов ресурс.

Поэтому одной из важнейших задач проектирования системы связи является такой выбор сигналов отдельных каналов, при котором взаимовлияние между каналами в идеале вообще бы отсутствовало; на практике же речь идет о том, чтобы оно не превосходило определенного допустимого уровня. Задачи разделения каналов, т.е. выделения сигналов интересующего нас канала из их смеси, проходящей через общий ресурс, решается в общем случае как приемными антеннами, так и различными устройствами последующего тракта приемника. В первом случае обработке подвергаются сигналы в виде электромагнитных волн, которые могут быть описаны векторными функциями от трех пространственных переменных и времени. Такие сигналы далее будем кратко именовать *векторными*. Начиная с выхода приемной антенны, имеем уже дело с сигналами, представляющими собой скалярные функции только времени. Такие сигналы ниже именуются *скалярными*. Соответственно в зависимости от обрабатываемых сигналов возможные методы разделения также подразделяются на скалярные и векторные.

Из векторных методов почти с самого начала развития радиотехники используется разделение сигналов, связанное с географическим положением различных терминалов. Сигналы от далеко отстоящих терминалов в силу быстрого роста затухания с увеличением расстояния не могут мешать работе данного терминала как бы ни были выбраны их параметры. Возможно разделять сигналы по направлению прихода соответствующей электромагнитной волны: например, антенна достаточно высокой направленности и с достаточно низким уровнем боковых лепестков позволяет выделить излучение, принятое с данного направле-

ния на фоне излучений, поступающих с иных направлений. Указанные методы, основанные на разделении излучений, в зависимости от географического положения их источников ниже именуется *разделением по пространству*. Другой вариант векторного разделения базируется на поляризационных свойствах излучений: два ортогональных по поляризации излучения могут быть эффективно разделены антеннами с соответствующими и достаточно хорошими поляризационными характеристиками.

Из изложенного ясно, что возможности векторного разделения не зависят от конкретных форм используемых радиосигналов. Таким образом, постановка задачи векторного разделения оказывается достаточно ясной: речь идет о том, чтобы на выходе приемной антенны присутствовал только сигнал нужного канала, а сигналы других каналов, в идеале, вообще отсутствовали.

Что касается скалярного разделения то в общем случае постановка задачи оказывается более сложной.

Прежде всего, сигналы всегда являются для получателя случайными процессами, так как их параметры никогда в точности не бывают заранее известны на приемном конце. Это относится к таким параметрам, как средняя амплитуда, начальная фаза, несущая частота, временные положения тактовых точек и т.п. (параметры переносчика). Тем более неизвестны значения передаваемого сообщения, а следовательно, и выбираемые при модуляции соответствующие сигналы. Таким образом, сигналы каждого (j -го) канала оказываются зависящими от двух групп параметров: параметров переносчика λ_j и модуляционных параметров ν_j . Параметры первой группы изменяются весьма медленно, так что для наших целей их можно считать постоянными. Наоборот, модуляционные параметры отображают символы передаваемого сообщения и потому изменяются оперативно. Итак, сигнал j -го канала ($j = 0, 1, 2, \dots, N$) можно записать в виде $s_j(\lambda_j, \nu_j, t)$. Далее, без ограничения общности можно считать канал с $j = 0$ выделяемым. Условно будем именовать его также полезным в отличие от остальных каналов, выступающих по отношению к выделяемому, как мешающие. Последние назовем также *соседними*.

Параметры переносчика выделяемого канала всегда определяются на приемном конце. Без этого невозможна демодуляция данного сигнала. Ниже при рассмотрении процедур разделения будем считать их известными. Для сигналов соседних каналов значения набора параметров λ_j остаются в точности неизвестными. На приемном конце известны лишь области их возможных значений.

Как отмечалось выше, в задаче разделения сигналов предполагается, что все они проходят через общую среду (используют общий ресурс). Примем, что в ней они суммируются с различными весами a_j , отображающими различие уровней сигналов и коэффициентов передачи трасс, по которым они поступают в эту среду. Суммарный сигнал, который ниже будем именовать *групповым*, принимает вид

$$U(t) = \sum_{j=0}^N a_j s_j(\lambda_j, \nu_j, t). \quad (6.1)$$

Очевидно, что полезный сигнал можно считать разделимым с сигналами других каналов только в случае, если он может быть однозначно определен по групповому сигналу, или, иначе говоря, если путем анализа этого группового сигнала можно всегда однозначно определить величину a_0 . Такой анализ должен проводиться на некотором временном интервале, содержащем некоторое число m тактов выделяемого сигнала. Поскольку положение тактовых точек полезного сигнала известно, можно принять без ограничения общности интервал анализа от $t = 0$ до $t = mT$. Пусть за это время транслируется m бит полезного сообщения, так что параметр ν принимает 2^m значений. Полагают, что полезный сигнал потен-

циально выделим из группового, если значение уровня a_0 однозначно определяется при любом v_0 .

По поводу сформулированной задачи разделения сигналов заметим следующее: хотя здесь явная процедура разделения не определялась, уже из самой постановки задачи следует, что соответствующие методы могут быть только линейными, поскольку измеряемый параметр (a_0) линейно связан с групповым сигналом (см. (6.1)). Линейность процедуры разделения очень важна в данном случае по следующей причине: реально групповой сигнал всегда наблюдается на фоне шумов и, если процедура разделения была бы нелинейной, то этот факт обязательно должен был бы учитываться. Учет шумов резко усложнил бы рассмотрение и врядли позволил бы довести его до конца. С другой стороны, при наличии достаточно большого числа сигналов соседних каналов с суммарным уровнем, значительно превосходящим полезный сигнал, нелинейные процедуры разделения наверняка оказались бы неэффективными, поскольку неизбежные комбинационные составляющие, вызванные взаимодействием сигналов соседних станций и шума, резко затруднили бы прием полезного сигнала. По этим соображениям приведенная выше постановка задачи разделения сигналов, исключая необходимость рассмотрения нелинейных процедур, которые, резко усложняя рассмотрение, все равно не привели бы к полезным результатам, является целесообразной и обоснованной.

6.1.2. Условия разделимости сигналов

Пусть имеется семейство функций $u_i(t)$, $i = 1, 2, \dots, N$. Некоторая функция $u_0(t)$, не равная тождественно нулю, называется *линейно независимой* от этого семейства, если не существует такого набора вещественных чисел b_i , что

$$u_0(t) = \sum_{i=1}^N b_i u_i(t). \quad (6.2)$$

Если фигурирующие в данных определениях функции зависят от некоторых параметров, причем сформулированное условие линейной независимости выполняется при любых значениях параметров каждой из функций, то говорят о *линейной независимости в усиленном смысле*.

Сформулируем и основное утверждение о разделимости сигналов, первоначальная формулировка которого принадлежит Д.В. Агееву: «...для того, чтобы полезный сигнал мог быть выделен из группового сигнала необходимо и достаточно, чтобы он был линейно независим в усиленном смысле от ансамбля сигналов соседних станций».

Перейдем к доказательству приведенного утверждения. Прежде всего, отметим, что поскольку групповой сигнал сформирован из сигналов соседних каналов и полезного сигнала в соответствии с (6.1), какое-нибудь решение для a_0 (возможно равное нулю) всегда существует. Таким образом, утверждение должно выявить условие, при котором это решение единственно. Проведем доказательство от противного. Пусть оно не единственно, так что одновременно

$$u(t) = a_0 s_0(\lambda_0, v_0, t) + \sum_{j=1}^N a_j s_j(\lambda_j, v_j, t) \text{ и } u(t) = d_0 s_0(\lambda_0, v_0, t) + \sum_{j=1}^N d_j s_j(\lambda_j, v_j, t),$$

причем $a_0 \neq d_0$. Отсюда

$$(a_0 - d_0) s_0(\lambda_0, v_0, t) = - \sum_{j=1}^N (a_j - d_j) s_j(\lambda_j, v_j, t),$$

что противоречит исходному предположению о независимости s_0 от семейства s_1, \dots, s_N . Таким образом, из независимости действительно следует однозначность определения a_0 , что и доказывает достаточность условия рассматриваемого утверждения.

Необходимость данного условия следует из того, что если бы s_0 было бы зависимо от семейства s_1, \dots, s_N , то выражалось бы рядом $s_0(\lambda_0, \nu_0, t) = \sum_{j=1}^N c_j s_j(\lambda_j, \nu_j, t)$, где c_j — некоторые вещественные числа. Поэтому величину $w(t) = s_0(\lambda_0, \nu_0, t) - \sum_{j=1}^N c_j s_j(\lambda_j, \nu_j, t) = 0$ можно

с любым коэффициентом прибавить к правой части (6.1). Таким образом, значение a_0 не только не определяется однозначно, но вообще может быть сделано любым. Это и доказывает необходимость сформулированного условия.

Развитые выше положения определяют интерес техники связи к ансамблям линейно независимых функций. Особенно часто используются подмножество таких ансамблей, именуемое *ортogonalными функциями*. Как уже отмечалось в гл. 3, две отличные от нулевых функции $u_0(t)$ и $u_1(t)$ называются *ортogonalными* на интервале от t_1 до t_2 , если выполняется соотношение

$$\int_{t_1}^{t_2} u_0(t)u_1(t)dt = 0. \tag{6.3}$$

Если эти функции зависят от некоторых параметров, и (6.3) выполняется при любых их значениях, то говорят об *ортogonalности в усиленном смысле*. Покажем теперь, что из ортogonalности следует линейная независимость. Действительно, пусть $u_0(t)$ ортogonalна каждой функции семейства $u_j(t)$ ($j = 1, 2, \dots, n$). Предположим, что она линейно зависима от такого семейства. Тогда существуют коэффициенты b_j , не все равные нулю, такие что выполняется (6.2). Помножая правую и левую части этого равенства на $u_0(t)$ и беря интеграл, найдем

$$\int_{t_1}^{t_2} u_0^2(t)dt = \sum_{i=1}^n b_i \int_{t_1}^{t_2} u_0(t)u_i(t)dt.$$

Последнее равенство невозможно: в его левой части стоит положительная величина (интеграл от квадрата ненулевой функции), а правая часть в силу ортogonalности функций тождественно равна нулю. Полученное противоречие и доказывает высказанное утверждение. Обратное утверждение о том, что из линейности следует ортogonalность, вообще говоря, неверно; функции могут быть линейно независимыми, но не ортogonalными. В определенном смысле, однако, линейно независимых функций столько же, сколько и ортogonalных, так что переход от линейно независимых сигналов к ортogonalным не сокращает возможностей выбора разделяемых сигналов. Вместе с тем для ортogonalных сигналов процедуры разделения оказываются более простыми, в связи с чем в технике связи, как правило, для трансляции различных каналов используются именно ортogonalные сигналы.

Отметим, что строгая ортogonalность на практике часто не достигается и приходится использовать квазиортogonalные сигналы. В связи с этим возникает задача оценки степени ортogonalности пары сигналов. Для ее разумного определения представим корреляцию между функциями в виде

$$\int_{t_1}^{t_2} u_0(t)u_1(t)dt = \frac{1}{\sqrt{E_0 E_1}} r, \tag{6.4}$$

где E_0 и E_1 — энергии функций $u_0(t)$ и $u_1(t)$ на интервале от t_1 до t_2 соответственно; r — некоторое число. Из известного неравенства Коши–Шварца следует, что определяемая этим соотношением величина r всегда удовлетворяет соотношению

$$|r| \leq 1. \quad (6.4a)$$

Величина r называется *нормированным коэффициентом корреляции сигналов*, ее модуль и может служить мерой их ортогональности: сигналы можно считать квазиортогональными, если $|r| \ll 1$.

Вернемся к общей постановке задачи о разделении сигналов и попытаемся более детально охарактеризовать цели соответствующих процедур. Как отмечалось, исходной здесь является необходимость исключения существенного влияния сигналов соседних каналов на прием полезного сигнала. В соответствии с материалами гл. 3 при приеме полезного сигнала осуществляются процедуры двух типов: 1) собственно демодуляция полезного радиосигнала, реализуемая, как правило, в соответствии с алгоритмами идеального приема (когерентного или некогерентного); 2) группа вспомогательных процедур (синхронизации), определяющих все необходимые параметры переносчика λ_j . Естественно, что сигналы соседних каналов не должны затруднять выполнение процедур обоих указанных типов.

Отметим, что если сигналы соседних каналов ортогональны в усиленном смысле полезному сигналу, то они теоретически совершенно не влияют на его демодуляцию идеальным приемником. В этом легко убедиться при рассмотрении структуры идеальных приемников (см. гл. 3). Действительно, все они начинаются с корреляторов, которые вообще не реагируют на входные сигналы ортогональные к полезному, так что защищать процесс демодуляции от таких соседних сигналов нет необходимости. Вместе с тем, как показано выше, использование в соседних каналах ортогональных в усиленном смысле сигналов реализует потенциальные возможности по числу одновременно функционирующих каналов. Итак, оказывается, что для всех практически интересных случаев в использовании специальных процедур разделения сигналов для защиты процессов демодуляции вообще нет необходимости.

Иначе обстоит дело с работой систем синхронизации приемников. На их функционирование сигналы соседних каналов могут оказывать существенное влияние. Поэтому часто важно защищать такие системы от сигналов соседних каналов, для чего и используются специальные системы разделения. Эти системы должны быть построены так, чтобы на их выходе полезный сигнал присутствовал бы практически в не искаженном виде, а сигналы соседних каналов были бы, по возможности, подавлены. Кроме того, необходимо чтобы не ухудшалось отношение сигнал/шум, что приводит, как указывалось выше, к требованию линейности соответствующих устройств.

6.1.3. Классификация ортогональных сигналов и способов разделения

Рассмотрим пару гармонических колебаний: $u_1(t) = \sqrt{2} \cos(\omega_1 t + \varphi)$ и $u_2(t) = \sqrt{2} \cos \omega_2 t$ каждое мощностью 1. Соответственно энергия этих сигналов на интервале времени от 0 до T составляет $E = T$. Имеем

$$r = \frac{1}{T} \int_0^T u_1(t) u_2(t) dt = \frac{\sin[(\omega_1 + \omega_2)T + \varphi] - \sin \varphi}{(\omega_1 + \omega_2)T} + \frac{\sin[(\omega_1 - \omega_2)T - \varphi] + \sin \varphi}{(\omega_1 - \omega_2)T}.$$

Строгие условия, при которых $r = 0$ таковы:

а) $(\omega_1 + \omega_2)T = 2\pi n$;

б) $(\omega_1 - \omega_2)T = 2\pi m$,

(6.5)

где m и n — целые числа. Условие «а» для радиосигналов несущественно, поскольку когда $(\omega_1 + \omega_2)T \gg |(\omega_1 - \omega_2)T|$, первым членом в (6.5) можно пренебречь, так что достаточным условием ортогональности оказывается выполнение требования «б».

Таким образом, при равенстве частот ($f_1 = f_2$) ортогональность ($r = 0$) достигается только при $\varphi = 0,5\pi$. Иначе говоря, синусоида и косинусоида одной и той же частоты (квадратурные сигналы) ортогональны на любом интервале T , причем не существует других гармонических колебаний той же частоты, им ортогональных. Пусть теперь частоты сигналов не равны. Тогда минимальный частотный интервал $\Delta f = |f_1 - f_2|$ между ними, при котором обеспечивается ортогональность, составляет $\Delta f = 1/T$. Далее при фиксированном, не равном 0, разnose между частотами, с ростом T коэффициент корреляции уменьшается, хотя и не монотонно, так что всегда можно выбрать столь большой интервал анализа T , что r по модулю будет сколь угодно мало. Таким образом, синусоиды разных частот f_1 и f_2 можно считать ортогональными при достаточно большом T (при $\Delta f T \gg 1$).

Сопоставим теперь эти данные с данными о спектре по Фурье отрезка синусоиды частоты f_c и длительности T . Легко вычислить, что спектр пропорционален величине

$$G(f) = \frac{\sin \pi(f - f_c)T}{\pi(f - f_c)T}. \quad (6.6)$$

Таким образом, если частотный интервал между отрезками синусоид $\Delta f = 1/T$, то их спектры существенно перекрываются, так что разделение этих сигналов путем фильтрации оказывается невозможным, хотя указанные сигналы, как показано выше, строго ортогональны и поэтому разделимы. Ниже ансамбль сигналов, состоящий из отрезков синусоид, следующих по частоте с интервалом $1/T$, будем именовать *гармоническим ансамблем с минимальным сдвигом* (ГАМС). Если разнос частот между сигналами $\Delta f = m/T$, где $m \gg 1$, то перекрытие их спектров достаточно мало, так что такие сигналы могут быть разделены посредством соответствующей фильтрации.

Разделение сигналов фильтрацией называется *разделением по частоте*, а соответствующий способ формирования группового сигнала — *частотным уплотнением*. Приведенные данные показывают, что разделение по частоте гармонических колебаний возможно, если произведение разноса частот на время анализа фильтра (длительность его импульсного отклика) много больше единицы. Поскольку радиосигналы всегда могут быть представлены в виде суммы достаточно большого числа гармонических составляющих, то можно утверждать, что любые сигналы с неперекрывающимися энергетическими спектрами допускают разделение по частоте с помощью соответствующим образом подобранных канальных фильтров. При этом чем меньше зазор между спектрами соседних сигналов, тем более крутые скаты должны иметь эти фильтры, т.е. тем большее время анализа приходится использовать.

При разделении по частоте перед подачей группового сигнала на входы систем синхронизации приемника из него путем фильтрации выделяется полезный сигнал. Важно отметить, что единственной характеристикой сигналов различных каналов, используемой при разделении по частоте, являются их энергетические спектры. Конкретные формы сигналов при разделении не используются и необходимы лишь при демодуляции.

В связи с понятием «разделение по частоте» сделаем следующие принципиальные замечания: *во-первых*, канальные сигналы никогда не имеют строго ограниченного спектра, так что энергетические спектры сигналов соседних каналов всегда пересекаются на том или ином уровне. Это обстоятельство является не только следствием конкретных условий реализации, но и носит принципиальный характер. Можно показать, что сигналы со строго ограниченным спектром не могут быть тождественно равны нулю ни на каком конечном вре-

менном интервале и, в частности, должны существовать всегда, нигде не начинаясь. Во-вторых, они могут быть предсказаны по любому своему отрезку на любое время вперед с любой требуемой точностью. Ясно, что все эти свойства не могут быть присущи сигналам — переносчикам информации. Поэтому идеальное разделение сигналов по частоте невозможно. На практике стремятся лишь так выполнить эту процедуру, чтобы взаимовлияние каналов имело бы допустимые величины, не ухудшающие существенно характеристик систем синхронизации.

Аналогичная ситуация имеет место относительно возможностей реализации канальных фильтров. Можно показать, что фильтры, коэффициент передачи которых на некотором конечном интервале частот тождественно равен нулю, являются физически не реализуемыми, так что речь может идти лишь о достаточно большом, а не полном подавлении сигналов соседних каналов с помощью реальных фильтров. Следует также отметить, что разделение близлежащих частот с помощью фильтров всегда приводит к появлению задержки сигнала и при том тем большей, чем большее подавление мешающего сигнала требуется. В связи с данными положениями достижимое качество частотного разделения каналов специально рассматривается в следующем разделе.

Сигналы с неперекрывающимися спектрами образуют важный класс ортогональных сигналов. Другой класс образуют сигналы, не перекрывающиеся по времени. Общая структура ансамбля таких сигналов может быть описана следующим образом: выбирается некоторый временной интервал, именуемый *кадром*, который разбивается на n непересекающихся подинтервалов — *слотов*. Слоты внутри кадра последовательно нумеруются (от 1 до n). Кадры периодически повторяются, причем j -е слоты каждого кадра отводятся для передачи сигналов j -го канала. Таким образом, информация каждого канала передается периодически следующими отрезками (пакетами). В данном случае сигналы каждого канала занимают свой слот, так что они не пересекаются во времени и поэтому являются ортогональными. Разделение каналов достигается здесь соответствующей коммутацией: вход приемника открывается только на время прохождения слот полезного сигнала. Такой способ разделения именуется *временным разделением*, а соответствующий способ формирования группового сигнала — *временным уплотнением*. При временном разделении требуется знание только временных положений полезных сигналов, другие их характеристики не требуются. Перед подачей на системы синхронизации приемника при временном уплотнении полезный сигнал выделяется из группового путем соответствующей коммутации.

Основные недостатки временного уплотнения связаны с необходимостью распределения информации каждого канала на передающей стороне по отдельным пакетам и объединения их на приемной стороне в непрерывный информационный поток без потерь и нарушений целостности сообщений. Для решения этих задач в технике связи разработаны специальные методы.

В начале 6.1.3 мы уже столкнулись с примером ортогональных сигналов, не допускающих разделение ни по частоте, ни по времени. Имелись в виду отрезки гармонических колебаний длительностью T с частотным разносом $\Delta f = 1/T$. Другим примером ортогональных функций не делимых ни по частоте, ни по времени является семейство функций В.А. Котельникова, для которых ниже используется обозначение $K_j(t)$ ($j = \dots -1, 0, 1, 2, \dots$)

$$K_j(t) = \frac{T \sin \pi \frac{t - jT}{T}}{\pi(t - jT)}, \quad (6.7)$$

где T — параметр функций, смысл которого ясен из рис. 6.1.

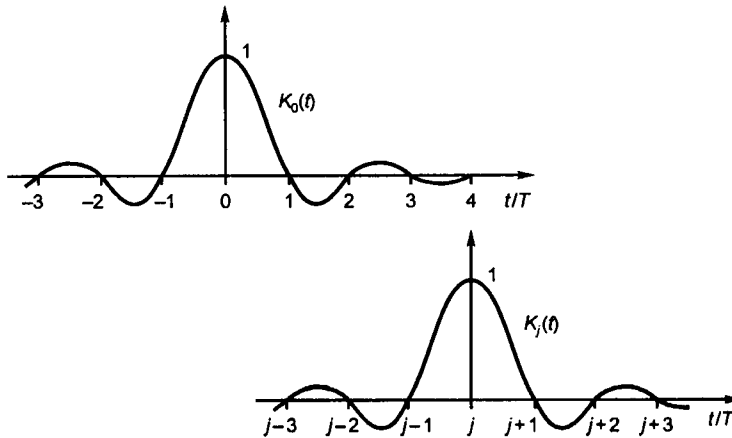


Рис. 6.1. Семейство функций В.А. Котельникова

Выражение (6.7) определяет семейство функций с бесконечным числом членов, причем отдельные функции отличаются друг от друга только временным сдвигом. Последнее означает, что модули спектра для всех функций совпадают, так что все они лежат в одной и той же полосе частот. Далее, они перекрываются и по времени, как это видно из рис. 6.1. Между тем на бесконечном временном интервале они ортогональны. Действительно, обозначая $x = \pi t/T$ и полагая $m \neq 0$, найдем

$$\begin{aligned} \int_{-\infty}^{\infty} K_0(t) K_m(t) dt &= \frac{\pi}{T} \int_{-\infty}^{\infty} \frac{\sin x \sin(x - m\pi)}{x(x - m\pi)} dx = \\ &= \frac{1}{mT} \int_{-\infty}^{\infty} \frac{\sin x \sin(x - m\pi)}{x} dx - \frac{1}{mT} \int_{-\infty}^{\infty} \frac{\sin x \sin(x - m\pi)}{(x - m\pi)} dx = 0. \end{aligned}$$

Последнее равенство следует из того, что последние два интеграла равны, в чем можно убедиться, произведя во втором из них замену: $y = m\pi - x$. Итак, функции Котельникова попарно ортогональны, хотя совмещены по спектру и пересекаются как функции времени. Можно показать, что модуль спектра функций Котельникова постоянен на интервале круговых частот от $-1/(2T)$ до $1/(2T)$ и равен нулю вне данного интервала. Таким образом, это видеосигналы. Однако на их основе легко сформировать и радиосигналы $U_m(t)$, осуществив преобразование на любую несущую частоту f в соответствии с соотношением $U_m(t) = K_m(t) \cos(2\pi f t + \varphi)$. Очевидно, что указанные выше свойства ортогональности функций Котельникова сохраняются.

Приведенные примеры показывают, что функции, не пересекающиеся по спектру и времени, далеко не исчерпывают все возможные варианты ортогональных сигналов, так что за пределами двух отмеченных выше классов функций (ортогональных по частоте и времени) остается еще ряд ансамблей ортогональных функций. Все эти ансамбли объединяют в класс функций ортогональных по форме. Соответственно к методам разделения по частоте и времени добавляется еще и разделение по форме, которое иногда именуют *кодовым разделением*.

Среди функций ортогональных по форме значительное место занимают группы ансамблей, являющихся пограничными с классами функций ортогональных по частоте и по времени. К ним относятся приведенные выше в качестве примеров ансамбли ортогональных функций. Так, ГАМС можно рассматривать как набор сигналов, отличающихся по частоте,

но на величину недостаточную для их эффективного частотного разделения. Впрочем, в тех случаях, когда разделение используется для защиты систем синхронизации, часто достаточным оказывается качество разделения, при котором существенная часть мощности близлежащих соседних сигналов накладывается на выделяемый полезный сигнал. Здесь к ГАМС может быть применено частотное разделение. Аналогично ансамбль функций Котельникова можно рассматривать как набор функций отличающихся по времени, скорость спадания которых вне выделенного для них временного интервала недостаточна для их эффективного разделения по времени. Тем не менее и здесь таким способом может быть достигнуто разделение функций, правда весьма грубое.

Существуют ортогональные по форме сигналы, которые полностью перекрываются как по частоте, так и по времени, вследствие чего их никак нельзя отнести к отмеченным выше пограничным группам. Такими сигналами являются так называемые *широкополосные сигналы* (ШПС), для которых характерна сравнительно сложная форма элементарных посылок. Именно, если при обычных, или *узкополосных сигналах* (УПС) элементарные посылки представляют собой отрезки гармонических колебаний с постоянными или сравнительно медленно изменяющимися параметрами, то при ШПС за время элементарной посылки T_s эти параметры многократно значительно изменяются по тому или иному достаточно сложному закону. В результате при УПС основная полоса энергетического спектра сигнала имеет величину порядка $\Delta f = 1/T_s$, а при ШПС она значительно расширяется, так что может быть представлена в виде $\Delta f = B/T_s$, где $B \gg 1$. Величина B именуется *базой* ШПС. Структура и свойства ШПС будут подробно рассматриваться в гл. 7. Здесь мы лишь проиллюстрируем особенности разделения таких сигналов по форме. Рассмотрим один из вариантов широкополосных сигналов, построенных путем дополнительной фазовой манипуляции (ФМ ШПС).

При формировании таких сигналов вначале реализуется один из обычных способов цифровой манипуляции — АМ, ЧМ, ФМ или их модификации. Соответствующие радиосигналы обозначим через $v_x(t)$. Далее спектр этих сигналов расширяют путем их фазовой манипуляции одной или несколькими маркерными цифровыми последовательностями, длительность символов которых τ_0 такова, что $T_s/\tau_0 = B \gg 1$. Последовательности названы здесь *маркерными*, поскольку их вид не зависит от передаваемой информации и заранее известен на приемной стороне радиолинии (заложен в приемник). Часто маркерные последовательности по своей структуре выбираются аналогичными случайному телеграфному сигналу, в связи с чем их называют также *псевдослучайными последовательностями* (ПСП). Чаще всего используют двоичные периодические ПСП с достаточно большим числом символов в периоде; в качестве символов ПСП принимают 1 и -1 . В качестве типовой двоичной ПСП можно принять последовательность, получающуюся в результате повторения независимых статистических испытаний с двумя равновероятными исходами, например подбрасывания монеты. Если сопоставить две такие последовательности, полученные в независимых сериях опытов, то можно утверждать, что вероятность совпадения исходов опытов в обеих последовательностях на каждом такте составляет 0,5, так что данные последовательности будут квазиортогональными. Отметим, что типовая ПСП оказываются квазиортогональной не только по отношению другим ПСП, но и по отношению к собственным сдвигам на один или большее число символов. Обозначим ПСП как функцию времени через $q(t)$.

Обычный метод формирования ФМ ШПС $v_{ш}(t)$ описывается соотношением $v_{ш}(t) = q(t)v_x(t)$. Умножение узкополосного манипулированного сигнала на ПСП означает его дополнительную двоичную фазовую манипуляцию. В результате спектр сигнала расширяется до величины $\Delta f = 1/\tau_0 = B/T_s$, т.е. ориентировочно, в B раз.

Может показаться, что использование ШПС сопряжено с резким снижением эффективности использования полосы частот (в B раз). Между тем это не так. Оказывается, что по

мере расширения Δf соответственно увеличивается и число ортогональных ШПС, которые могут размещаться в этой полосе. Каждый из таких сигналов занимает всю полосу и передается непрерывно и тем не менее они попарно ортогональны. В соответствии с вышеизложенным такие сигналы могут быть разделены по форме. Последнее позволяет по мере расширения полосы, располагать в ней все большее число каналов, так что резкого снижения эффективности использования полосы не происходит. Более детальный анализ показывает, что эффективность использования полосы за счет ШПС может не только не ухудшаться, но и для некоторых условий связи — увеличиваться. На этом факте основано использование систем сотовой связи с разделением сигналов по форме (стандарт IS 95, см. гл. 11).

Сигналы отдельных каналов при использовании ФМ ШПС и разделении по форме строятся следующим образом: передающие тракты всех терминалов системы формируют свои узкополосные модулированные сигналы $v_{y,j}(t)$ ($j = 0, 1, 2, \dots, n - 1$, где n — число каналов) на одной и той же несущей частоте. Эти сигналы дополнительно помножаются в трактах передачи на ПСП, причем для каждого канала используется своя ПСП $q_j(t)$. Эти ПСП должны быть попарно ортогональны при любом их взаимном сдвиге, если такой сдвиг может возникать в канале связи. И так, излучаемые ФМ ШПС сигналы имеют вид

$$v_{\omega,j}(t) = q_j(t)v_{y,j}(t). \tag{6.8}$$

При приеме полезного сигнала используется сформированная в приемнике копия $g_0(t)$ той ПСП $q_0(t)$, с помощью которой на передающей стороне было осуществлено расширение его спектра. Поскольку в системах связи, как правило, не используется единое время, да и запаздывание при распространении сигнала от передатчика до приемника неизвестно, $g_0(t)$ оказывается сдвинутым по времени относительно $q_0(t)$ на заранее неизвестную величину. Поэтому в приемнике используется специальная процедура — поиск и синхронизации по задержке, обеспечивающая совмещение по времени ПСП, наложенной на принимаемый полезный сигнал, и ее местной копии. В результате чего

$$g_0(t) = q_0(t). \tag{6.9}$$

Соответствующий алгоритм выделения полезного сигнала по форме представлен блок-схемой на рис. 6.2.

На вход этой схемы поступает групповой сигнал $u(t)$, в который входит и полезный сигнал $v_{\omega,0}(t) = q_0(t)v_{y,0}(t)$. После входного перемножителя этой схемы полезный сигнал принимает вид $g_0(t)q_0(t)v_{y,0}(t) = q_0^2(t)v_{y,0}(t) = v_{y,0}(t)$, так как $q_0^2(t) = 1$. Иначе говоря, в результате линейной операции, выполняемой перемножителем, дополнительная фазовая манипуляция, произведенная при формировании ШПС, устраняется и восстанавливается узкополосный полезный сигнал. При этом полоса частот, занимаемых полезным сигналом, сокращается в B раз, вследствие чего указанную процедуру называют *сверткой по частоте*. На рис. 6.3 схематически представлен энергетический спектр полезного сигнала до свертки $S_{\omega}(f)$ и после нее $S_y(f)$.

После свертки узкополосный полезный сигнал целиком пройдет через полосовой фильтр и поступит на вход демодулятора, который должен выполнять его демодуляцию.

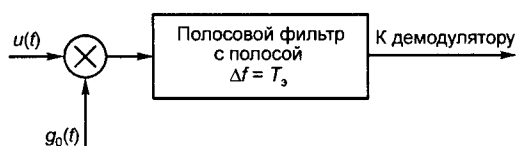


Рис. 6.2. Блок-схема выделения ФМ ШПС

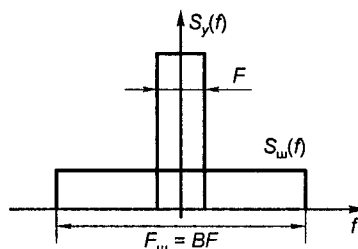


Рис. 6.3. Преобразование энергетических спектров сигналов при свертке по частоте

Остановимся теперь на прохождении через приведенную схему сигналов соседних каналов. Если бы $g_0(t)$ была бы строго ортогональной любой из ПСП этих сигналов $q_j(t)$, то на выходе схемы на рис. 6.2 эффект от их воздействия полностью бы отсутствовал. Строгая ортогональность требует жесткой синхронизации ПСП различных сигналов, что удается реализовать далеко не всегда. В типовых случаях, когда между ПСП сигналов отдельных каналов имеются случайные сдвиги по времени, в устройстве (см. рис. 6.3) просто не происходит свертки по частоте сигналов соседних каналов, и полосы занимаемых ими частот не изменятся, так что их энергетический спектр и на выходе устройства останется приблизительно равным $S_{ш}(f)$. В результате на выход полосового фильтра пройдет лишь небольшая часть $(1/B)$ мощности сигнала соседнего канала. Таким образом, свертка по частоте повышает отношение мощности полезного сигнала к мощности помех от соседней станции в B раз. Если обозначить мощность j -го сигнала через P_j , то отношение мощности полезного сигнала на входе демодулятора к мощности помех от соседних станций на том же входе составит

$$H = \frac{P_0 B}{\sum_{j=1}^{n-1} P_j} = \frac{B}{\sum_{j=1}^{n-1} H_j^{-1}}, \quad (6.10)$$

где $H_j = P_0/P_j$. В частном случае, когда мощности сигналов всех каналов одинаковы ($H_j = 1$), получим

$$H = B/(n - 1). \quad (6.11)$$

Обратим внимание читателя на то, что описанный алгоритм разделения ШПС по форме не требует точного знания параметров полезного сигнала. Так, этот алгоритм не использует манипуляционные параметры, вследствие чего процедуры выделения сигнала и его демодуляции здесь разделены, нет необходимости знать заранее фазу несущей, а ее частота должна быть известна лишь со сравнительно низкой точностью, соизмеримой с полосой сигнала до расширения его спектра.

Описанный алгоритм выделения ФМ ШПС является типовым для широкополосных сигналов. При их разделении всегда вначале используется линейное устройство, преобразующее полезный сигнал в узкополосный (устройство свертки по частоте), после чего для разделения применяются уже методы, характерные для УПС и, прежде всего, разделение по частоте или времени.

В заключение отметим, что на практике реализуется комбинация различных методов разделения. При этом скалярное разделение почти всегда дополняется теми или иными способами векторного разделения. Скалярных методов применяется сразу несколько. Часто используется частотное уплотнение, хотя бы уже в силу того, что для различных систем выделяются, как правило, различные полосы рабочих частот. Но и в пределах одной системы общая полоса делится на ряд частотных интервалов. Внутри каждого из них может также применяться частотное уплотнение, но может иметь место и временное деление или разделение по форме. В действительности здесь речь идет о последовательной реализации на передающей и приемной сторонах нескольких методов уплотнения (разделения).

Так, в системах сотовой связи стандарта GSM реализуется вначале временное, а затем частотное уплотнение. На первом этапе уплотнения группа терминалов формирует на одной и той же (по номиналу) несущей частоте пакеты своих сообщений и при том так, чтобы пакеты этой группы не перекрывались по времени. Все такие пакеты излучаются в эфир, образуя так называемую «спицу» — последовательность не перекрывающихся по времени пакетов на одной и той же (по номиналу) несущей частоте. Другим группам терминалов выделяются другие несущие частоты, расстройки между которыми выбраны так, чтобы спектры соответствующих «спиц» не пересекались. Таким образом, между «спицами» реализуется частотное разделение.

В системах сотовой связи стандарта IS 95 реализуется сочетание частотного деления и разделения по форме. В некоторой полосе на общей (по номиналу) несущей частоте передается группа ФМ ШПС (уплотнение по форме). Таких непересекающихся полос в системе используется несколько (уплотнение по частоте). При разделении вначале должно осуществляться разделение по частоте, а затем по форме.

Возможен и иной вариант сочетания частотного уплотнения и уплотнения по форме, отличающийся от описанного порядком использования этих методов. Здесь сначала формируется ряд узкополосных манипулированных сигналов на различных несущих частотах, выбранных так, чтобы спектры сигналов не перекрывались (частотное уплотнение). Затем каждый из данных сигналов дополнительно манипулируется ПСП (общей для всех сигналов или индивидуальными), вследствие чего спектры сигналов расширяются в B раз и начинают существенно перекрываться. Для разделения образовавшегося группового сигнала должен вначале применяться описанный выше алгоритм свертки по частоте, а затем частотное разделение.

6.1.4. Многостанционный доступ при сотовой связи

Выше мы уже несколько раз упоминали системы сотовой связи. Подчеркнем теперь, что главным признаком, позволяющим выделить эти системы в отдельный класс, является сочетание двух наиболее важных групп методов уплотнения: векторных (разделение по пространству) и различных скалярных. Для наземных систем сотовой связи характерно использование пространственного разделения, связанного с различием дальности связи, хотя в некоторых вариантах таких систем используется и формирование сот, основанное на различии углов прихода сигналов. Что касается спутниковых систем сотовой связи, то здесь всегда используется разделение именно по направлению прихода сигналов. Для пояснения сотового принципа организации связи остановимся на некоторых определениях.

Будем именовать *областью обслуживания* (ОО) системы радиосвязи тот географический регион или район, где размещаются радиостанции (РС), функционирующие в данной системе. Различают радиолинии и каналы связи, понимая под *радиолинией* тракт между двумя РС — приемной и передающей. В отличие от этого под каналом связи понимают путь, по которому транслируется данная конкретная информация в рассматриваемой системе радиосвязи. В общем случае канал состоит из нескольких включенных последовательно одна за другой радиолиний.

В полносвязных системах реализуется возможность организации радиолиний между любой парой РС. В ряде случаев используются системы ограниченной связности. Например, часто реализуются радиальные системы, в которых имеется центральная радиостанция (ЦРС) и организуются радиолинии только между нею и остальными РС. Иногда создаются трехуровневые системы, в которых функционируют станции трех уровней: оконечные (ОРС), узловые или базовые (БРС) и центральная (ЦРС) станции. Каждая БРС поддерживает радиолинии только с определенной подгруппой ОРС, другими БРС и ЦРС. В этой схеме БРС и ЦРС используются как средства коммутации и ретрансляции принимаемых каналов, что и обеспечивает возможность создания каналов связи между любыми станциями сети. Используются и системы без ЦРС; тогда связи между БРС организуются не по радио, а с помощью проводных или оптоволоконных средств. Часто БРС и ЦРС данной системы радиосвязи сопрягают с различными другими системами связи, действующими на территории ее ОО или охватывающими эту область, что позволяет создавать составные каналы связи нескольких систем.

При создании систем радиосвязи, как правило, оказывается полезным не стремиться обеспечивать радиоканалы между любой парой РС независимо от их размещения в ОО,

а разбить эту область на ряд зон (сот) и создавать радиолинии только внутри таких сот (зональное или сотовое обслуживание). При этом в каждой соте создается БРС, поддерживающая радиальную связь со всеми ОРС этой соты. Тем или иным способом обеспечивается также связь БРС различных сот между собой, что и обеспечивает возможность создания каналов связи между любой парой РС системы. Системы, использующие разбиение ОО на соты, именуются *сотовыми*, в отличие от интегральных систем, не использующих этот принцип.

Сотовые системы по сравнению с интегральными обеспечивают значительные выигрыши как по требуемому энергетическому потенциалу радиолиний, так и по использованию полосы рабочих частот системы радиосвязи. Энергетический выигрыш определяется двумя факторами: сокращением дальности связи и более рациональным распределением транслируемой информации по ОО.

Первый фактор очевиден, особенно если учесть, что для наземных систем радиосвязи интегральное обслуживание не слишком малых областей связано с загоризонтным распространением радиоволн, что всегда сопряжено со значительными энергетическими потерями. Второй фактор менее очевиден, хотя и не менее значим. Для его пояснения целесообразно разделить информацию, которая может транслироваться в системах связи, на индивидуальную и групповую (радиальную). К первому типу относится информация, предназначенная только одной РС, и не представляющая интереса для всех остальных станций системы. Примером может служить телефонная информация. В отличие от этого групповая информация представляет интерес для группы станций или даже для всех станций соты или системы (циркулярная информация). Примером может служить радиовещательная информация. При интегральном обслуживании такое различие типов передаваемой информации не учитывается и сигналы, отображающие всю циркулирующую в системе информацию, транслируются на всю ОО. В сотовых системах передача информации может дифференцироваться, так чтобы на каждую соту поступала только та информация, которая предназначена РС этой соты. Ясно, что такая дифференциация позволяет экономить энергетические ресурсы системы связи.

Возможность повышения эффективности использования полосы частот при зональном обслуживании определяется возможностью повторения частот при передаче сигналов в различных сотах. Такая возможность возникает в достаточно разнесенных по пространству сотах, так что уровень сигналов каждой из них, достигающих другой соты, оказывается относительно малым.

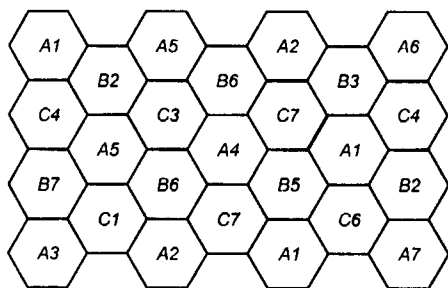


Рис. 6.4. Гексагональная структура сот

Размещение сот внутри ОО существенно зависит от местных условий и может быть достаточно разнообразным. В качестве основной, хотя и грубой, модели такого размещения можно принять гексагональную сетку (рис. 6.4).

Здесь каждая сота граничит с шестью другими сотами. Ясно, что в соседних сотах должны использоваться только линейно разделимые сигналы. В противном случае по крайней мере в районах, примыкающих к границам сот, сигналы отдельных РС будут существенно мешать друг другу, так что их прием станет невозможным. По этой причине для соседних сот приходится использовать различные группы сигналов, такие, что они оказываются линейно разделимыми. В соответствии с вышеизложенным данные группы могут различаться, например, по поляризации, частоте, времени или форме.

Как следует из рис. 6.4, использование трех групп таких разделимых сигналов (групп А, В и С) позволяет при гексагональном размещении сот исключить работу соседних сот на

сигналах одной и той же группы. Однако сигналы данной соты создают значительные помехи не только в соседних сотах, но и в сотах, окружающих соседние. Поэтому приходится исключать использование одной и той же группы сигналов во всех сотах, граничащих с любой данной. Оказывается, что минимальное число групп сигналов, необходимое для реализации этого требования, — семь. На рис. 6.4 показано требуемое размещение по сотам семи групп сигналов (1, 2, ..., 7). Действительно, какую бы соту на рисунке мы не рассмотрели, окажется, что все шесть граничащих с нею сот имеют различные группы сигналов, при том отличные от ее собственной группы. Решение этой задачи при числе групп сигналов менее семи невозможно. В сравнительно редких случаях, чтобы еще более разнести по пространству соты с повторяющимися сигналами, прибегают к использованию девяти групп сигналов.

Оценим теперь общее число линейно делимых сигналов, которые необходимы для всей ОО при сотовом обслуживании. Для простоты примем, что по трафику соты идентичны, так что для обслуживания каждой из них необходимо одно и то же число сигналов v . Тогда общее число требуемых сигналов составит $N = mv$, где m — число используемых групп сигналов.

Остановимся на вопросе о том, какое число сигналов потребовалось бы для той же ОО при интегральном обслуживании. Здесь следует различать случаи индивидуальной и групповой информации. При передаче *индивидуальной информации* общее число радиолиний с ОРС при переходе от интегрального обслуживания к сотовому не меняется: оно остается равным числу ОРС, одновременно работающих во всей ОО. Поэтому при индивидуальной информации и сотовом обслуживании при M сотах достигается выигрыш в общем числе требуемых в системе сигналов по сравнению с вариантом интегрального обслуживания в

$$n = M / m \text{ раз.} \quad (6.12)$$

Поскольку число сот может достигать нескольких сотен, то последнее соотношение означает, что требуемое число сигналов сокращается в десятки раз. Если, например, эти сигналы создаются на основе разделения по частоте, то это соответствует экономии спектра рабочих частот, необходимого для функционирования системы, в то же число раз.

Иначе обстоит дело при трансляции в системе циркулярной информации. При интегральном обслуживании требуется организация одной радиолинии на все ОРС, размещенные в ОО, в то время как при зональном обслуживании должны создаваться M таких радиолиний (по одной в каждой соте). Соответственно требуемое число делимых сигналов возрастет при переходе к сотовому обслуживанию в m раз (при $M > m$).

Подводя итог, можно сказать, что при трансляции индивидуальной информации сотовое обслуживание позволяет обеспечить резкое повышение эффективности использования связанных ресурсов всех видов и поэтому является основным направлением дальнейшего совершенствования соответствующих систем связи. При трансляции циркулярной информации целесообразность использования сотового обслуживания зависит от конкретных условий и требует каждый раз специального анализа.

6.2. Эффективность использования полосы рабочих частот в системах с разделением по частоте

6.2.1. Внеполосные излучения при базовых методах манипуляции

Как уже отмечалось, радиосигналы с цифровой модуляцией представляют собой случайные процессы. Поэтому занимаемая ими полоса частот характеризуется их энергетическим спектром. Этот спектр концентрируется обычно вокруг несущей частоты сигнала, причем подавляю-

щая часть мощности сигнала приходится на некоторую ограниченную полосу частот F . Сама полоса чаще всего определяется шириной энергетического спектра, измеренной по уровню минус 3 дБ относительно максимума, и именуется *основной полосой сигнала* (ОПС). Составляющие спектра, лежащие вне ОПС, именуются *внеполосными излучениями* (ВИ). Такие составляющие существуют всегда, поскольку, как отмечалось выше, спектр реального сигнала не может быть равен нулю ни на каком конечном частотном интервале. Уровень спектральных составляющих ВИ с ростом отстройки f их частоты от несущей всегда убывает, хотя и немонотонно.

Существенную роль в приеме того сигнала, которому они принадлежат, ВИ не играют, так как по уровню спектра они близки или даже существенно ниже спектральной плотности мощности шума, на фоне которого осуществляется этот прием. Вместе с тем ВИ могут создавать помехи приему сигналов соседних станций, которые граничат по спектру с полезным сигналом. Приходится увеличивать разнос между частотами несущих колебаний соседних станций, что приводит к ухудшению использования полосы рабочих частот. Этим обстоятельством и определяется стремление максимально ослабить ВИ манипулированных радиосигналов. Методам решения данной задачи и посвящен настоящий раздел.

Количественная оценка эффективности подавления ВИ основывается на показателе достижимой за счет этого удельной скорости передачи информации γ . При этом γ определяется (см. 5.1) с учетом защитных интервалов между частотами соседних несущих, необходимых для ослабления влияния ВИ до допустимого уровня. Конкретная методика вычисления γ сводится к следующему:

- задаются допустимым увеличением спектральной плотности помех приему полезного сигнала, вызванным наложением на флуктуационные шумы ВИ соседних каналов; соответствующее значение потерь выбирается исходя из того, насколько напряженной является энергетика радиолиний рассматриваемой системы связи;
- исходя из энергетических спектров используемых сигналов, требуемых скоростей передачи и мощностей канальных сигналов определяется частотный план системы радиосвязи, учитывающий и требуемые защитные интервалы между частотами несущих отдельных каналов;
- определяется удельная скорость передачи информации $\gamma = C_{\Sigma}/F_{\Sigma}$, бит/с/Гц, где C_{Σ} — суммарный объем сообщений, передаваемых по всем каналам системы связи в единицу времени; F_{Σ} — общая полоса рабочих частот, занимаемых системой связи.

Как было показано в предыдущих главах, для одноканальных систем требования полосной и энергетической эффективности оказываются, как правило, противоречивыми: повышение полосной эффективности обычно приводит к снижению энергетической, и наоборот. Аналогичная ситуация имеет место и в многоканальных системах. Соответственно, рассматривая методы подавления внеполосных излучений, необходимо стремиться, в первую очередь, использовать те из них, которые либо вообще не увеличивают требуемые значения h_b , либо увеличивают их в допустимой степени.

После общих замечаний перейдем к конкретному анализу полосной эффективности различных способов манипуляции. Начнем с так называемых классических, или базовых методов манипуляции (см. 1.3.4), к которым мы относим амплитудную, частотную и фазовую со скачкообразными изменениями манипулируемого параметра. Начнем с анализа энергетических спектров соответствующих радиосигналов. Для чего остановимся сначала на решении вспомогательной задачи — определении энергетического спектра случайной телеграфной последовательности (СТП). Под СТП понимают цифровой случайный видеосигнал, принимающий на каждом тактовом интервале длительностью T равновероятно одно из двух возможных значений $+1$ или -1 (символы на различных тактовых интервалах независимы).

По общему правилу для определения энергетического спектра $S_{\text{СТП}}(f)$ найдем функцию корреляции процесса, а затем произведем над ней преобразование Фурье. На рис. 6.5 представлена возможная реализация СТП, а ниже та же реализация, но запаздывающая на время τ , причем $\tau < T$. В заштрихованных областях значения процессов всегда совпадают, поскольку принадлежат одной и той же посылке. На остальных (незаштрихованных) интервалах вероятность совпадения значений этих функций, так же как и их несовпадения, равна 0,5. Поэтому среднее значение их произведения на данных интервалах равно нулю, т.е. функция корреляции СТП есть

$$R(\tau) = \frac{T - \tau}{T} = 1 - \frac{\tau}{T}. \quad (6.13)$$

При $\tau > T$ любая посылка СТП $v(t)$ не пересекается с нею же в сдвинутом сигнале $v(t - \tau)$, так что $R(\tau) = 0$. Вид функции корреляции СТП представлен на рис. 6.6, а.

Беря от этой функции преобразование Фурье, найдем энергетический спектр

$$\begin{aligned} S_{\text{СТП}}(f) &= \int_{-T}^T R(\tau) \cos(2\pi f \tau) d\tau = \\ &= T \frac{\sin^2(\pi f T)}{(\pi f T)^2}, \end{aligned} \quad (6.14)$$

график которого представлен на рис. 6.6, б.

Как видно из рис. 6.6, максимум спектра, равный $S_m = T$, приходится на нулевую частоту, а основная полоса СТП лежит в интервале от данной частоты до $f_{\text{он}} = 0,5/T$. Выше по частоте располагаются ВИ. Огибающая их спектра для $f > f_{\text{он}}$ может быть представлена в виде

$$Q(f) = \frac{S_m}{\pi^2 v^2} \text{ при } v \geq 0,5, \quad (6.15)$$

где v — обозначена обобщенная расстройка, в данном случае $v = fT$. Выражение (6.15) показывает, что при увеличении обобщенной расстройки в два раза («на октаву») огибающая спектра ВИ снижается в четыре раза (на 6 дБ). Таким образом, скорость убывания спектра СТП составляет 6 дБ на октаву. Можно показать, что именно такая скорость присуща всем случайным процессам, реализации которых имеют разрывы непрерывности первого рода (т.е. ограниченные по величине скачкообразные изменения значения). Далее, общая теория показывает, что если у реализаций некоторого случайного процесса при любом t имеются

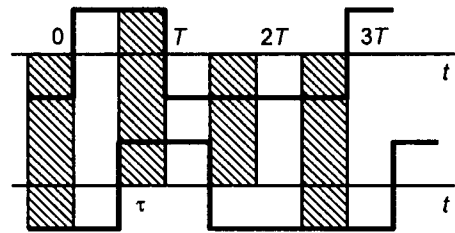


Рис. 6.5. К определению функции корреляции СТП

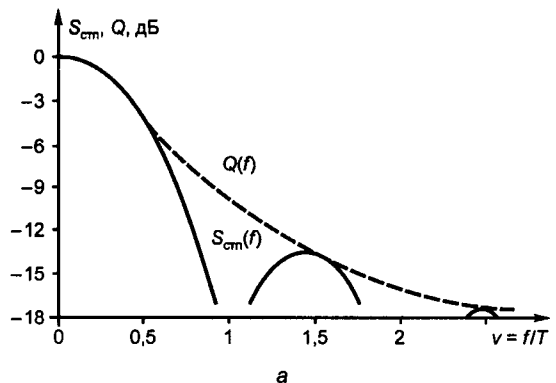
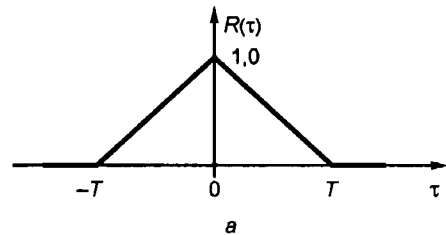


Рис. 6.6. Коэффициент корреляции СТП (а) и энергетический спектр СТП (б)

производные до $b - 1$ порядка включительно, то скорость убывания его энергетического спектра составляет $6b$ дБ на октаву. Поэтому, если сглаживать закон изменения манипулируемого параметра, то можно достичь снижения уровня ВИ. Именно этот путь, как мы увидим ниже, и составляет основу одного из наиболее распространенных в настоящее время способов решения данной задачи.

Перейдем к оценке энергетических спектров радиосигналов с цифровой модуляцией. Такие сигналы могут быть получены путем преобразования СТП на частоту несущей f_n . Так, радиосигнал ФМ2 мощностью P_c можно представить в виде

$$V(t) = \sqrt{2P_c} v(t) \cos(2\pi f_n t), \quad (6.16)$$

а сигнал АМ при средней мощности P_c в виде

$$V(t) = \sqrt{P_c} [1 + v(t)] \cos(2\pi f_n t) = \sqrt{P_c} \cos(2\pi f_n t) + \sqrt{P_c} v(t) \cos(2\pi f_n t). \quad (6.17)$$

Из выражения (6.16) следует, что спектр ФМ2 может быть получен из спектра СТП в виде

$$S_{\text{ФМ2}} = 0,5 P_c S_{\text{СТП}}(|f - f_n|). \quad (6.18)$$

Отсюда следует, что основная полоса энергетического спектра ФМ2 составляет

$$F_{\text{оп}} = 1/T_s. \quad (6.18a)$$

Для огибающей спектра внеполосных составляющих ФМ2 остается справедливым соотношение (6.15) при соответствующем определении обобщенной расстройки: $\nu = (|f - f_n|)T_s$, и значения максимума спектра, который для сигнала ФМ2 составляет, как это следует из (6.18), $S_m = P_c T_s$.

Отметим теперь, что энергетический спектр сигнала ФМ4 совпадает с таковым для ФМ2 при том же значении T_s . Действительно, сигнал ФМ4 при использовании манипуляционного кода Грея (см. гл. 5) можно рассматривать как сумму двух каналов ФМ2, образованных с использованием квадратурных несущих. Ясно, что сигналы этих каналов ФМ2 ортогональны и поэтому их энергетические спектры суммируются. Мощность суммарного сигнала равна сумме мощностей сигналов каждого канала, откуда следует, что при равных длительностях элементарных сигналов энергетические спектры ФМ4 и ФМ2 совпадают. Следует иметь в виду, что при том же T_s битовая скорость ФМ4 в два раза превосходит битовую скорость ФМ2, что при сохранении скорости кода, соответствует в два раза большей скорости передачи информации. Таким образом, при передаче информации с одинаковой скоростью и при том же кодировании ширина спектра ФМ4 оказывается на любом уровне в два раза уже таковой для ФМ2, что соответствует удвоению удельной скорости передачи информации. Напомним, что, как показано в гл. 3, энергетическая эффективность обоих способов манипуляции (параметр h_6) одинакова. Таким образом, переход от ФМ2 к ФМ4 повышает в два раза полосную эффективность системы без снижения ее энергетической эффективности. При дальнейшем увеличении кратности манипуляции k полосная эффективность продолжает повышаться пропорционально k . Однако начинает и быстро снижаться энергетическая эффективность (повышаться требуемое значение h_6 , см. гл. 5).

Переходя к спектру сигнала АМ, отметим, что, как это следует из (6.17), он состоит из неслучайного гармонического колебания несущей частоты, вообще не отображающего манипулирующий сигнал и отбирающего половину средней мощности сигнала, и манипулированного ФМ2 колебания той же мощности, собственно и переносящего информацию. Это позволяет очевидным образом выразить спектр АМ через только что определенный спектр ФМ2. Отметим также, что наличие в спектре АМ неманипулированной несущей объясняет причину, по которой, как было показано в гл. 3, потенциальная помехоустойчивость АМ на 3 дБ ниже, чем у ФМ2 или ФМ4. Напомним, что приведенные рассуждения относятся к

бинарной АМ. Можно показать, что для k -кратной АМ при том же значении T_3 спектр практически не меняется. Если же сохранять постоянной скорость передачи информации при том же коде, спектр будет сокращаться пропорционально k , а потенциальная помехоустойчивость снижаться ориентировочно на

$$\eta \text{ [дБ]} = 20 \lg(2^k - 1) - 10 \lg k. \quad (6.19)$$

При амплитудно-фазовой манипуляции (КАМ, см. гл. 5) спектральная эффективность повышается пропорционально росту длительности элементарного сигнала T_3 , но существенно снижается энергетическая эффективность.

Что касается спектра сигналов ЧМ, то при достаточно больших разносах несущих частот элементарных сигналов (обобщенная расстройка больше 1) эти спектры в первом приближении можно оценивать путем суммирования спектров АМ сигналов с пассивной паузой на каждой из этих несущих.

С помощью полученных данных оценим достижимую удельную скорость передачи сообщений при использовании классической ФМ2. Оценим допустимый разнос несущих частот соседних каналов при многоканальной передаче сигналов ФМ2 посредством частотного разделения. Положим вначале, что скорости передачи и мощности для всех сигналов совпадают. Пусть эти сигналы принимаются на фоне белого гауссовского шума (БГШ) со спектральной плотностью N_0 . Средний уровень энергетического спектра каждого сигнала в его основной полосе, как показано выше, составляет ориентировочно $w_c = P_c T_3$. Отношение этой плотности к N_0 для уверенного приема, как это следует из теории потенциальной помехоустойчивости, должно составлять примерно 10, так что по порядку величины $N_0 = 0,1 P_c T_3$.

Для того чтобы влияние ВИ соседних каналов не слишком сказывалось на помехоустойчивости приема сигнала данного канала, их энергетический спектр должен быть достаточно малым по сравнению с N_0 , например, не превосходить $0,1 N_0$. Такая оценка получена для равных по мощности сигналов различных каналов. На практике это требование не выполняется. Действительно, в различных каналах могут использоваться приемные системы различной чувствительности (отличающиеся, например, размерами антенн), что требует и использования сигналов с различными мощностями. Влияют также разбросы реальных параметров приемо-передающих трактов, даже когда их номинальные значения одинаковы. Условно примем, что с учетом этих обстоятельств необходимо еще на порядок снизить допустимый уровень спектра ВИ, так чтобы он не превосходил $0,01 N_0 = 0,001 P_c T_3 = 0,001 S_m$. Исходя из данного требования найдем, используя (6.15), что разнос несущих частот должен быть не менее

$$F_p = 10/T_3. \quad (6.20)$$

Последнее соотношение позволяет определить удельную скорость передачи информации при классической ФМ2:

$$\gamma_k \approx 0,1. \quad (6.21)$$

Рассуждая аналогичным образом, для ФМ4 найдем

$$\gamma_k \approx 0,2. \quad (6.22)$$

6.2.2. Подавление внеполосных излучений при ФМ путем сглаживания скачков фазы

Как показано в 6.2.1, ВИ, присущие классическим методам манипуляции, резко снижают их полосную эффективность. В связи с этим на первый план выходит задача подавления ВИ. Ниже рассмотрены пути решения задачи применительно к энергетически наиболее эффек-

тивному из базовых методов манипуляции — фазовой. Реализация соответствующих мер приводит к возникновению специфических вариантов, которые можно рассматривать как разновидности ФМ либо как новые способы манипуляции.

Сглаживание закона изменения манипулируемого параметра является одним из основных путей ослабления ВИ. Применительно к ФМ сглаживание означает замену скачкообразных изменений фазы несущей при манипуляции на ее плавное изменение. Главная проблема заключается в том, чтобы, с одной стороны, действительно обеспечить значительное снижение ВИ, а с другой — существенно не ухудшить потенциальную помехоустойчивость манипуляции. Последнее может явиться следствием изменения формы используемых элементарных сигналов, а следовательно, и величин коэффициентов корреляции между ними (см. гл. 3).

Охарактеризуем качественно указанные эффекты. Пусть интервал времени, на котором производится сглаживание закона изменения фазы, $t_c \leq T_s$. Как бы ни был мал данный интервал всегда можно так выбрать закон изменения фазы на нем, чтобы исключить разрывы как самой фазы, так и любого числа ее производных. Как уже отмечалось, это должно приводить к значительно более быстрому спадаанию огибающей спектра ВИ $Q(\nu)$. Следует иметь в виду, что этот эффект будет проявляться только при достаточно больших расстройках,

грубо говоря при $|f - f_n| > \frac{1}{t_c} = F_m \frac{T_s}{t_c}$, а при меньших расстройках спектр ВИ останется

тем же, что и без сглаживания. По этой причине необходимо стремиться к увеличению интервала сглаживания t_c и приближению его к длительности элементарного сигнала T_s .

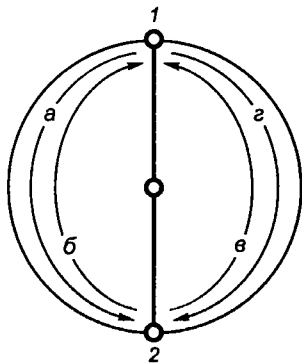


Рис. 6.7. Векторная диаграмма сигналов ФМ2 при сглаживании скачков фазы

Что касается самого закона сглаживания, то он может быть выбран различными способами. Наиболее простым считается линейное сглаживание, когда фаза изменяется от своего начального положения к конечному линейно. Однако возможны различные варианты. Чтобы пояснить их происхождение, обратимся к рис. 6.7, где представлены векторные диаграммы сигналов. Примем, что плоскость рисунка (фазовая плоскость) вращается с круговой частотой f_n , так что гармоническое колебание этой частоты представляется неподвижным вектором на этой плоскости. Тогда два возможных ЭС классической ФМ2 представляются векторами 1 и 2.

Для классической ФМ2 переход из состояния 1 в 2 и наоборот происходит скачком. При сглаживании для этих переходов могут использоваться различные интервалы времени и различные пути: возвращение из 2 в 1 может

происходить как по пути «б», так и по пути «в», а движение из 1 в 2 по путям «а» или «г». Выбор указанных путей может быть фиксированным или осуществляться в каждой тактовой точке по тому или иному закону. Таким образом, имеется ряд возможностей, приводящих к различным разновидностям ФМ2. Не следует думать, что различия между данными разновидностями незначительны. Сравним, например, систему, использующую пути «а» и «в», т.е. движение по всей фазовой плоскости с полным вращением (ПВ) вектора, и систему, использующую только пути «а» и «б», т.е. с неполным вращением (НВ) вектора сигнала. Обратим внимание на то обстоятельство, что линейное изменение фазы вектора соответствует сдвигу его несущей частоты.

Таким образом, в варианте ПВ используются две частоты: f_n (сохранение состояния 1 или 2) и $f_1 = f_n + 0,5/t_c$ (пути «а» и «в»), а в НВ — три: f_n ; $f_1 = f_n + 0,5/t_c$ (путь «а») и

$f_2 = f_n - 0,5/t_c$ (путь «б»). Соответственно отличаются и спектры манипулированных радиосигналов: если в варианте НВ спектр симметричен относительно частоты f_n , то в случае ПВ это не так. В случае ПВ удобно рассмотреть среднюю частоту $f_0 = 0,5(f_n + f_1)$. Если использовать фазовую плоскость, вращающуюся с этой круговой частотой (рис. 6.8), то на ней за время t_c изображающий вектор будет поворачиваться на 90° в ту или иную сторону (по часовой стрелке или против нее) в зависимости от того, какой именно символ сообщения (1 или 0) передается. В тактовых точках вектор принимает одно из положений 0, 90° , -90° или 180° .

6.2.3. Манипуляция с минимальным сдвигом

Остановимся специально на варианте с ПВ при максимальной длительности сглаживания $t_c = T_3$. Соответствующий метод манипуляции называется *манипуляцией с минимальным сдвигом* (ММС). Две используемые частоты имеют значения $f_1 = f_0 - 0,25/T_3$ и $f_2 = f_0 + 0,25/T_3$. Сама частота f_0 является виртуальной и в сигнале ММС не представлена. Реально используемые частоты сдвинуты относительно нее на $f_d = 0,25/T_3$ вверх или вниз. Разность частот $f_2 - f_1 = 0,5/T_3 = 2f_d$. В соответствии с (6.5) два радиоимпульса с частотами несущих колебаний f_1 и f_2 и длительностью T_3 ортогональны, причем при меньшем значении разности их частот ортогональность не может быть обеспечена. Поэтому именно это значение $2f_d$ и именуют *минимальным частотным сдвигом*, откуда и происходит название метода манипуляции. Таким образом, ММС может рассматриваться как двоичная частотная манипуляция с частотами f_1 и f_2 . Однако важная особенность такого способа манипуляции состоит в наличии жесткой связи девиации частоты и длительности передачи элементарных символов. Эта связь требует, чтобы набег фазы (относительно гармонического колебания с частотой f_0) составлял за время посылки точно плюс или минус 90° , а фазовая траектория в тактовые моменты была непрерывной. Поэтому данный метод манипуляции можно рассматривать и как разновидность фазовой модуляции при специальном способе сглаживания фазовых траекторий.

Соответствующая векторная диаграмма ММС изображена на рис. 6.8, где представлены четыре состояния (А, Б, В и Г), в которых может находиться вектор, отображающий сигнал ММС в моменты времени, совпадающие с тактовыми точками. Между ними этот вектор совершает круговое вращение в ту или в иную сторону, переходя из данного состояния в соседнее, причем *соседним* называется состояние, отличающееся от данного на 90° в ту или другую сторону. Например, ближайшими к Г являются состояния В и А. Примем в качестве положительного направления отсчета фазы — вращение против часовой стрелки. Тогда вращение в положительном направлении означает передачу частоты f_2 , а в отрицательном — частоты f_1 . Именно выбор в каждой тактовой точке направления вращения или подлежащей передаче на очередном тактовом интервале частоты и определяет значение очередного передаваемого бита. Соответствие между значением этого бита и направлением вращения вектора, именуемое *манипуляционным кодом*, может устанавливаться для ММС различными способами. Наиболее предпочтительный из них будет описан позднее.

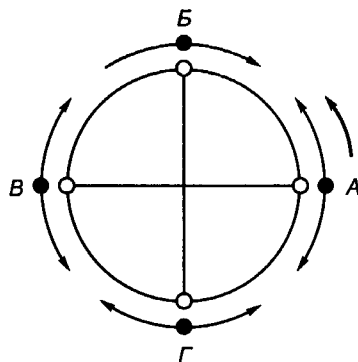


Рис. 6.8. Векторная диаграмма сигналов ММС

Соответствующая векторная диаграмма ММС изображена на рис. 6.8, где представлены четыре состояния (А, Б, В и Г), в которых может находиться вектор, отображающий сигнал ММС в моменты времени, совпадающие с тактовыми точками. Между ними этот вектор совершает круговое вращение в ту или в иную сторону, переходя из данного состояния в соседнее, причем *соседним* называется состояние, отличающееся от данного на 90° в ту или другую сторону. Например, ближайшими к Г являются состояния В и А. Примем в качестве положительного направления отсчета фазы — вращение против часовой стрелки. Тогда вращение в положительном направлении означает передачу частоты f_2 , а в отрицательном — частоты f_1 . Именно выбор в каждой тактовой точке направления вращения или подлежащей передаче на очередном тактовом интервале частоты и определяет значение очередного передаваемого бита. Соответствие между значением этого бита и направлением вращения вектора, именуемое *манипуляционным кодом*, может устанавливаться для ММС различными способами. Наиболее предпочтительный из них будет описан позднее.

Поскольку, как показано выше, ММС представляет собой некоторый специальный вариант двоичной ЧМ, может возникнуть вопрос, нельзя ли распространить на ММС метод идеального приема и данные о потенциальной помехоустойчивости, полученные для коге-

рентного приема двоичных сигналов в 3.2. Для того, чтобы понять, что такой подход был бы ошибочным, необходимо вспомнить условия применимости теории идеального приема: одно из важнейших предположений заключается в независимости сигналов, передаваемых на различных тактовых интервалах. Это предположение часто нарушается, когда используется сглаживание закона изменения фазы при манипуляции. Действительно, для достижения сглаживания необходимо, чтобы закон изменения фазы на данном тактовом интервале зависел не только от передаваемого на нем символа сообщения, но и от значения элементарного сигнала, переданного на предыдущем такте, ибо необходимо обеспечить непрерывность изменения фазы при переходе от предыдущего такта к данному. Такая зависимость имеет место и для ММС. По этой причине приведенные выше результаты теории идеального приема не могут быть непосредственно применены к методам манипуляции со сглаживанием манипулируемого параметра, включая ММС. Из нижеследующего станет ясно, что к ММС эта теория может быть все-таки применена, но некоторым специальным образом. Прежде более детально изучим структуру сигналов данного метода манипуляции.

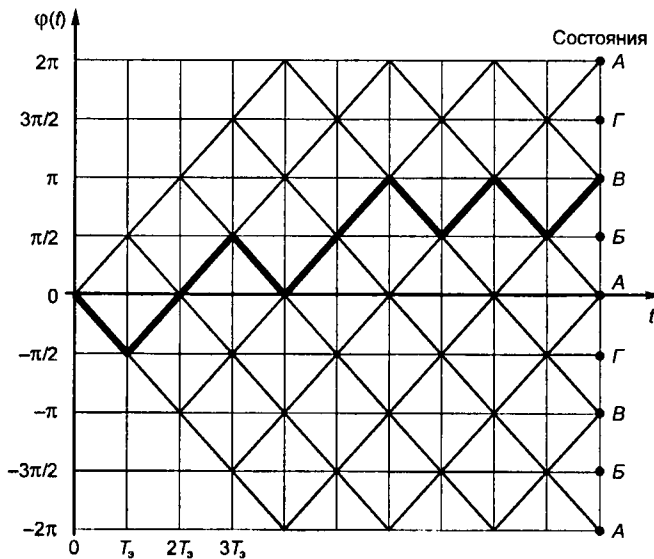


Рис. 6.9. Фазовая решетка при ММС

На рис. 6.9 представлены возможные фазовые траектории сигнала ММС, т.е. графики отклонения фазы сигнала от фазы гармонического колебания с частотой f_0 для различных возможных последовательностей передаваемых двоичных символов при условии, что начальное значение отклонения фазы принято равным нулю. Какая именно из представленных траекторий будет соответствовать той или иной последовательности манипулирующих бит, зависит от используемого манипуляционного кода. На рис. 6.9 одна из возможных фазовых траекторий выделена жирной линией. Фазовая решетка по-существу линейная развертка векторной круговой диаграммы, представленной на рис. 6.8. Последняя позволяет получить и другое полезное описание ММС сигналов — их квадратурное представление.

Рассмотрим проекции вектора, отображающего ММС сигнал, на оси абсцисс и ординат диаграммы рис. 6.8 — $Q(t)$ и $I(t)$ соответственно. Примем, что в начальный момент времени вектор занимал положение, соответствующее нулевой фазе. Возможные значения $I(t)$ представлены на рис. 6.10, а, $Q(t)$ на рис. 6.10, б. Графики этого рисунка построены в соответствии с фазовой решеткой, представленной на рис. 6.9. Легко видеть, что в зависимости от значений манипулирующих символов проекции принимают значения, изображенные сплошными или штриховыми линиями; переход со сплошных на штриховые линии может происходить только в точках перехода данной проекции через нуль. По форме эти проекции представляют собой полуволны синусоиды (косинусоиды) протяженностью $2T_s$. Из приведенных данных можно заключить, что фактически здесь имеет место двухканальная передача, причем каналы образованы на квадратурных несущих $\cos 2\pi f_0 t$ и $\sin 2\pi f_0 t$. Посылки на несущих имеют форму полуволны синусоиды длительностью $2T_s$, посылки в квадратурных каналах сдвинуты друг относительно друга на половину своей длительности. (Отметим, что методы манипуляции, при которых посылки в квадратурных каналах сдвинуты по времени друг относительно друга, именуется *офсетными*.)

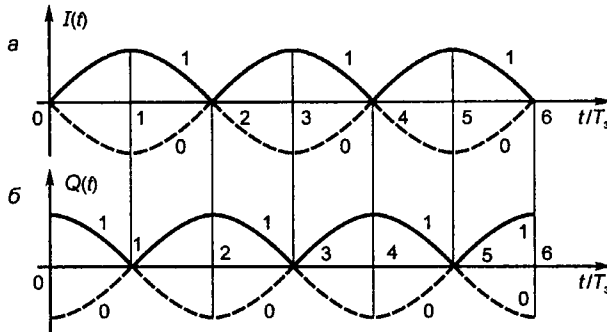


Рис. 6.10. Квадратурное представление сигнала ММС

Итак, сигнал ММС может быть представлен в виде

$$u(t) = I(t) \sin 2\pi f_0 t + Q(t) \cos 2\pi f_0 t. \tag{6.23}$$

Оба слагаемых правой части равенства (6.23) представляют собой манипулированные методом ФМ2 радиосигналы с квадратурными несущими с огибающими элементарных посылок, повторяющими сплошные кривые рис. 6.10, а и б соответственно. Иначе можно сказать, что используется два квадратурных радиосигнала-переносчика: $v_s(t) = |\sin(\pi t / T_s)| \sin(2\pi f_0 t)$ и $v_c(t) = |\cos(\pi t / T_s)| \cos(2\pi f_0 t)$ с длительностями элементарных посылок по $2T_s$; посылки различных каналов сдвинуты по времени на T_s . Переносчики модулируются по фазе методом ФМ2, так что после манипуляции фазы несущей в течение каждой из посылок принимают одно из двух возможных состояний (либо нуль, либо π) в зависимости от того, выбрано положительное или отрицательное значение амплитуды данной посылки на данной квадратуре (сплошная или штриховая кривая). Таким образом, каждая посылка длительностью $T_k = 2T_s$ на каждой из квадратур переносит ровно 1 бит. Легко видеть также, что огибающая суммарного сигнала $A(t) = \sqrt{I^2(t) + Q^2(t)} = \text{const}$, а средние мощности квадратурных радиосигналов $P_c = P_s = 0,5P_\Sigma$, где P_Σ — мощность суммарного сигнала. В соответствии с указанными значениями мощностей радиосигналов и длительностей элементарных сигналов мож-

но заключить, что значения h_6 для суммарного сигнала и сигналов отдельных подканалов одинаковы и равны $h_6^2 = P_c T_k = P_\Sigma T_\Sigma / N_0$.

Что касается связи между значениями фаз несущих в подканалах и значениями бит манипулирующей последовательности, то она определяется видом манипуляционного кода. Особенно простой эта связь оказывается при следующем манипуляционном коде: в состояниях A и Γ (см. рис. 6.8) положительное вращение вектора (использование частоты f_2) соответствует передаче бита со значением a_1 , а отрицательное его вращение (частота f_1) — бита со значением a_0 ; в состояниях B и B , наоборот, частота f_2 соответствует a_0 , а f_1 — a_1 . В такой записи под a_1 и a_0 можно понимать 1 и 0 соответственно, а можно, наоборот, 0 и 1. Такое использование частот приводит к следующему порядку передачи информации по квадратурным каналам: модулирующая последовательность бит разделяется на два потока, в первый из которых попадают биты, поступающие на нечетных тактовых интервалах (первый поступивший бит, третий, пятый и т.д.), а во второй на четных (втором, четвертом и т.д.). Далее биты первой последовательности манипулируют методом ФМ2 квадратуру с амплитудой — $Q(t)$, а второй — с $I(t)$. Соответствующая манипуляция осуществляется за счет выбора сплошной или штриховой линии на каждом временном интервале, протяженностью $2T_3$.

Таким образом, рассматриваемый манипуляционный код соответствует разделению транслируемого потока бит на два равноскоростных подпотока и передаче каждого из них на своей квадратуре. На приемной стороне демодулированные подпотоки объединяются в общий поток путем поочередного извлечения бит из подпотоков. Использование других манипуляционных кодов приводило бы к необходимости более сложной процедуры объединения на выходе демодулятора подпотоков в общий поток, при которой каждый бит общего потока формировался бы путем той или иной логической операции над несколькими демодулированными битами подпотоков. Поскольку всегда есть определенная вероятность ошибки в демодулированном бите, такая процедура приводила бы к размножению ошибок. В связи с тем, что описанный манипуляционный код от этого недостатка свободен, он используется на практике. Ниже будем исходить именно из применения такого манипуляционного кода. В качестве примера отметим, что выделенная фазовая траектория (см. рис. 6.9) соответствует при таком манипуляционном коде следующей последовательности передаваемых бит: 0, 1, 1, 1, 1, 0, 1, 0, 1, 0.

На рис. 6.11, *a* представлена структурная схема модулятора ММС, реализующего описанный выше алгоритм манипуляции. Двоичное цифровое сообщение с тактовой частотой $f_T = 1/T_3$, выделяемой системой тактовой синхронизации (СТС), разделяется вначале на два подпотока (четных и нечетных бит). Разделение осуществляется с помощью переключателя П (контакты «ч» для выхода четных бит и «н» — для нечетных), работой которого управляет выделенная СТС тактовая частота. Подпотоки поступают каждый на свой расширитель (P_c и P_s соответственно), увеличивающий длительность каждого бита до $2T_3$; тактовая частота подпоследовательностей становится равной $f_T/2$, а сами подпоследовательности — офсетными. Схема содержит генератор квадратурных переносчиков (ГКП) $v_s(t)$ и $v_c(t)$; длительности формируемых посылок синхронизованы с тактовой частотой f_T , чем обеспечивается ее точное равенство $2T_3$. В умножителях переносчики манипулируются подпоследовательностями по фазе по методу ФМ2. Выходной радиосигнал ММС формируется суммированием квадратурных каналов в соответствии с (6.23).

Переходя к вопросу о потенциальной помехоустойчивости ММС, рассмотрим приемник такого сигнала, состоящий из двух параллельно включенных приемников отдельных подканалов (рис. 6.11, *b*).

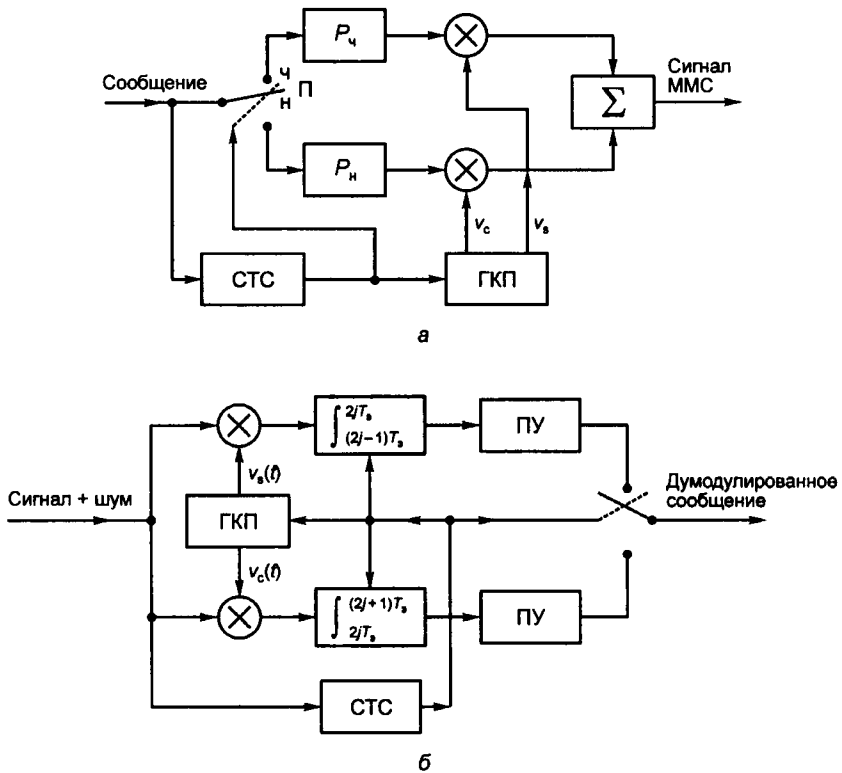


Рис. 6.11. Структурная схема модулятора (а) и демодулятора (б) ММС

Поскольку в подканалах используются сигналы ФМ2, то идеальные когерентный приемники для них (см. 3.2) имеют вид, представленный на рис. 6.11, б. Отметим, что у интеграторов демодуляторов подканалов время интегрирования $2T_s$ (синхронизуется от СТС приемника). Аббревиатурой ПУ обозначены пороговые устройства с нулевым порогом, аналогичные реализуемым в общей схеме идеального приемника ФМ (см. рис. 3.5). Канальные приемники начинаются с корреляторов, так что приемник одного подканала не будет реагировать на сигнал другого в силу ортогональности квадратурных сигналов. Далее сигнал одного подканала не может также способствовать приему другого подканала, поскольку по подканалам транслируются независимые сообщения. Это позволяет заключить, что рассматриваемый приемник ММС сигналов является идеальным. Вероятность ошибки в каждом подканале, а следовательно, и для всего сигнала ММС, определится стандартным выражением для фазовой телеграфии

$$P_{\text{ош}} = \Phi(\sqrt{2}h_s), \tag{6.24}$$

так что потенциальная помехоустойчивость ММС, такая же как и у ФМ2 и не снизилась из-за проведенного предельного сглаживания скачков фазы ФМ сигналов.

Остановимся еще на одном возможном варианте построения модулятора ММС, основанном на трактовке данного метода манипуляции как некоторой специальной разновидности двоичной ЧМ. Можно было бы просто использовать ЧМ-манипулятор, ставя в соответствие значениям транслируемых бит 0 и 1 определенные значения частот (например, 0 —

частоту $f_1, 1 - f_2$). При этом для того, чтобы получить ММС необходимо, чтобы разность этих частот была бы равна $0,5/T_6$, а фазы ЭС соседних тактовых интервалов в тактовой точке совпадали. Однако (см. рис. 6.8) такая манипуляция приводила бы к манипуляционному коду, отличному от желаемого, так что потребовалось бы перекодирование демодулированных бит. Чтобы сохранить принятый манипуляционный код, необходимо перед ЧМ-модулятором осуществить специальное перекодирование транслируемого сообщения. Из анализа диаграммы на рис. 6.8 следует, что алгоритм такого перекодирования должен соответствовать соотношению

$$x_k = b_{2j-1} \oplus b_{2j} \text{ при } k = 2j \text{ и } x_k = b_{2j} \oplus b_{2j+1} \text{ при } k = 2j+1, \quad (6.25)$$

где b_j — двоичная информационная последовательность, подаваемая на вход преобразователя; x_k — двоичная последовательность на его выходе; \oplus — суммирование по mod 2. Последовательность x_k подается на модулятор ЧМ2, т.е. каждое значение ее символа соответствует одной из двух возможных частот, выдаваемых этим модулятором.

Структурная схема преобразователя, соответствующего (6.25), представлена на рис. 6.12. Штриховой линией ограничен регистр сдвига с отводами, работа которого тактируется с тактовой частотой поступающего двоичного сообщения. Символом \bar{x} обозначен инвертор.

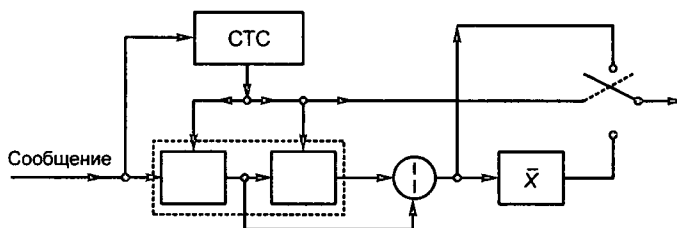


Рис. 6.12. Структурная схема преобразователя

Перейдем к характеристике спектральной эффективности ММС. Как показывают соответствующие расчеты, энергетический спектр сигналов ММС имеет вид

$$S(\nu) = \frac{8P_{\Sigma}T_6}{\pi^2} \frac{(1 + \cos 4\pi\nu)}{(1 - 16\nu^2)^2}, \quad (6.26)$$

где обобщенная расстройка $\nu = (f - f_0)T_6$.

На рис. 6.13 представлены графики энергетических спектров ММС сигнала, построенные в соответствии с (6.26), и сигналов ФМ2 и ФМ4, построенные по формуле (6.18). Для удобства сопоставления спектров они представлены как функции стандартной расстройки $\nu = (f - f_0)T_6$, где f_0 — средняя частота спектра для ФМ, совпадающая с частотой несущего колебания; T_6 — длительность двоичного символа сообщения.

Основная полоса частот для ММС составляет $F_{\text{оп}} = 0,6T_6$, т.е. несколько превосходит ее значение для ФМ4 ($F_{\text{оп}} \approx 0,5/T_6$). Зато спектр внеполосных излучений убывает у ММС, как и следовало ожидать, гораздо быстрее: со скоростью 12 дБ на октаву. Если определять размещение соседнего канала требованием подавления внеполосных излучений на 30 дБ, то в соответствии с данными рис. 6.13 положение его несущей частоты должно определяться значением $\nu = 1,2$, так что для ММС $\gamma = 1,2$, что почти на порядок выше, чем для ФМ4.

Таким образом, предельное сглаживание скачков фазы и соответственно переход к ММС позволяют резко повысить спектральную эффективность фазовой манипуляции без снижения ее помехоустойчивости.

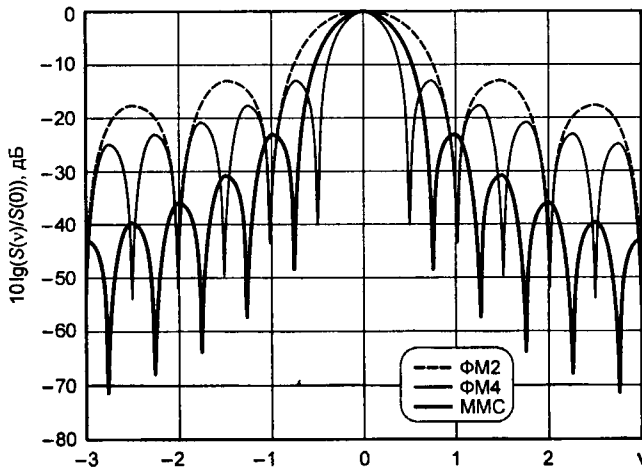


Рис. 6.13. Энергетические спектры сигналов ФМ2, ФМ4 и ММС

Описанный вариант ММС обеспечивает непрерывное изменение фазы при манипуляции, так что сигнал, рассматриваемый как функция времени, оказывается непрерывным. Поскольку закон изменения фазы имеет в тактовых точках изломы, то первая производная фазы в таких точках не существует. Более сложные законы сглаживания могли бы обеспечить не только непрерывность, но и наличие производных более высоких порядков, что в определенной степени улучшило бы спектр сигнала. Один из вариантов такого сглаживания рассматривается в следующем пункте.

6.2.4. Гауссовская манипуляция с минимальным сдвигом

Предположим, что реализуется способ формирования сигнала ММС с частотной модуляцией, для чего вначале осуществляется преобразование модулирующего сообщения в соответствии с (6.25) устройством рис. 6.12. Выходная последовательность символов преобразователя b_j представляется в виде биполярной функции времени, принимающей для $b_j = 1$ на тактовом интервале значение 1, а для $b_j = 0$ значение -1 . Скачки этой функции в тактовых точках привели бы при ММС к скачкообразным изменениям мгновенной частоты и соответственно к изломам в фазовых траекториях. Чтобы сгладить скачки модулирующей функции, ее пропускают через соответствующий фильтр нижних частот (ФНЧ). Такой метод сглаживания фазовых траекторий предусматривается, например, стандартами GSM-900, DCS-1800, DECT. Соответствующий способ модуляции при таком способе формирования радиосигнала принято называть *гауссовской модуляцией с минимальным сдвигом* (ГММС). Гауссовской эта модуляция именуется в связи с тем, что в качестве ФНЧ здесь используют фильтр с частотной характеристикой вида $\exp(-cf^2)$, который принято называть гауссовским.

Поскольку на выходе ФНЧ модулирующая функция уже не является двухуровневой, а принимает и промежуточные значения, то при манипуляции уже нельзя ограничиться формированием только двух частот, так что обычный модулятор ЧМ2 становится непригодным. Требуемый закон изменения частоты сигнала создают с помощью фазовой модуляции. Для этого вначале напряжение, отражающее требуемый закон изменения частоты, пропускают через идеальный интегратор, преобразовывая его таким образом в закон изменения фазы $\varphi(t)$, а его, в свою

редь, реализуют, суммируя квадратурные векторы с уровнями $\sin\varphi(t)$ и $\cos\varphi(t)$, соответственно. Структурная схема модулятора, построенного на этих принципах, представлена на рис. 6.14.

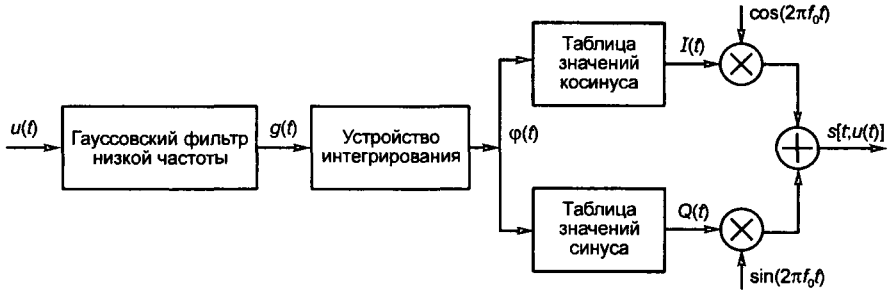


Рис. 6.14. Структурная схема модулятора ГММС

На вход модулятора подается преобразованный модулирующий сигнал

$$u(t) = \sum_{i=-\infty}^{\infty} x_i v(t - iT_s), \quad i = \dots, -1, 0, +1, \dots, \quad (6.27)$$

где $x_i \in \{-1, +1\}$ — информационный символ, передаваемый на интервале времени $iT_s < t \leq (i+1)T_s$, с номером i , T_s — длительность информационного символа, функция

$$v(t) = \begin{cases} 1/T_s, & \text{для } 0 < t \leq T_s, \\ 0, & \text{вне этого интервала} \end{cases} \quad (6.28)$$

определяет форму этого символа. Сигнал (6.27) подвергается предварительной низкочастотной фильтрации с помощью гауссовского фильтра низкой частоты, который должен иметь импульсный отклик следующей формы:

$$g(t) = B \sqrt{\frac{2\pi}{\ln 2}} \exp\left\{-\frac{2\pi^2 B^2 (t-1,5T_s)^2}{\ln 2}\right\} = \frac{1}{\sqrt{2\pi}\sigma_g} \exp\left\{-\frac{(t-1,5T_s)^2}{2\sigma_g^2}\right\}, \quad \sigma_g = \frac{\sqrt{\ln 2}}{2\pi B}, \quad (6.29)$$

где B — ширина полосы пропускания этого фильтра на уровне 0,707 от максимального значения его амплитудно-частотной характеристики

$$G(f) = \exp\left\{-\left(\frac{f}{B}\right)^2 \frac{\ln 2}{2}\right\}. \quad (6.30)$$

Из выражения (6.29) следует, что физически реализуемый фильтр не может иметь такой импульсный отклик, поскольку его значения при отрицательном значении аргумента t отличны от нуля, т.е. отклик фильтра появляется на выходе фильтра раньше начала воздействия на входе. Тем не менее можно считать, что такой фильтр может быть реализован приближенно, т.е. импульсный отклик физически реализуемого фильтра достаточно хорошо можно аппроксимировать функцией $g(t)$ для $t \geq 0$.

При дальнейшем описании данного сигнала более удобно использовать нормированное время $\tau = t/T_s$, при котором из (6.29) получаем

* Эти символы представляют собой результат предварительного кодирования сообщения в соответствии с (6.25).

$$g(\tau) = \frac{1}{\sqrt{2\pi}\sigma_g} \exp\left\{-\frac{(\tau-1,5)^2}{2\sigma_g^2}\right\}, \quad \sigma_g = \frac{\sqrt{\ln 2}}{2\pi BT_s}, \quad -\infty < \tau < +\infty. \quad (6.31)$$

Единственным параметром такого представления является произведение BT_s полосы пропускания фильтра на длительность прямоугольного импульса на его входе. Длительность информационного символа в стандарте GSM-900 принята $T_s = 3,69$ мкс и не может быть изменена. Значение полосы пропускания фильтра B может быть выбрано из разных соображений. Стандартом GSM-900 рекомендуется значение произведения $BT_s = 0,3$. Для стандарта DECT принято $BT_s = 0,5$. График функции (6.31) при $BT_s = 0,3$ изображен на рис. 6.15.

Отклик гауссовского фильтра на воздействие в виде одиночного прямоугольного импульса (6.28) вычисляется как свертка этих импульсов и отклика (6.29):

$$q(\tau) = \int_{-\infty}^{\tau} v(t)g(\tau-t)dt = \frac{1}{2} \left\{ \operatorname{erf} \left[\frac{\pi\sqrt{2}}{\sqrt{\ln 2}} (\tau-1,0) BT_s \right] - \operatorname{erf} \left[\frac{\pi\sqrt{2}}{\sqrt{\ln 2}} (\tau-2,0) BT_s \right] \right\}, \quad \tau \geq 0. \quad (6.32)$$

График отклика представлен на рис. 6.16. Важно здесь обратить внимание, что одиночный прямоугольный импульс на входе фильтра имеет длительность T_s , а отклик на выходе фильтра заметно отличен от нуля на интервале $3T_s$. В результате соседние символы на выходе гауссовского фильтра существенно перекрываются, т.е. имеет место существенная межсимвольная интерференция, или искажения (МСИ).

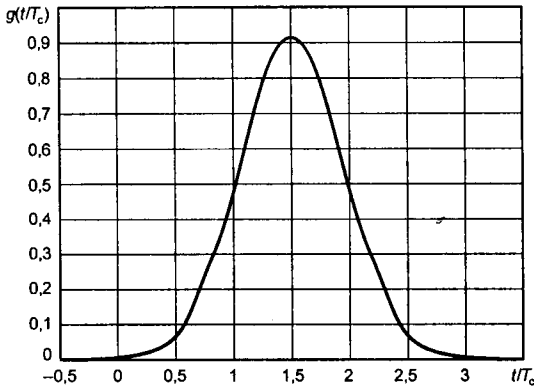


Рис. 6.15. Импульсный отклик гауссовского фильтра низкой частоты

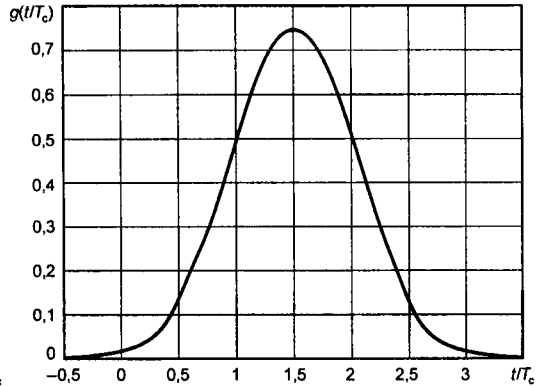


Рис. 6.16. Отклик гауссовского фильтра на одиночный прямоугольный импульс длительностью T_s

Теперь очевидно, что если модулирующий сигнал (6.27) подать на вход гауссовского фильтра низкой частоты, то сигнал на его выходе можно представить как сумму свертков (6.32):

$$p(\tau) = \sum_{i=-\infty}^{\infty} x_i q(\tau-i), \quad -\infty < \tau < +\infty. \quad (6.33)$$

Следовательно, сигнал на выходе устройства интегрирования на бесконечном интервале времени можно записать в виде следующей суммы:

$$\varphi(t) = \int_{-\infty}^t p(\tau) d\tau = \sum_{i=-\infty}^{\infty} x_i \int_{-\infty}^t q(\tau-i) d\tau = \sum_{i=-\infty}^{\infty} x_i \psi(t-i), \quad -\infty < t < +\infty. \quad (6.34)$$

Здесь введено обозначение

$$\psi(t) = \int_{-\infty}^t q(\tau) d\tau, \quad -\infty < t < +\infty, \quad (6.35)$$

для интеграла от одиночного отклика гауссовского фильтра, который называют *фазовой функцией*. График последней для случая, когда функция $q(\tau)$ имеет вид, представленный на рис. 6.16, изображен на рис. 6.17. Значения фазовой функции изменяются только на интер-

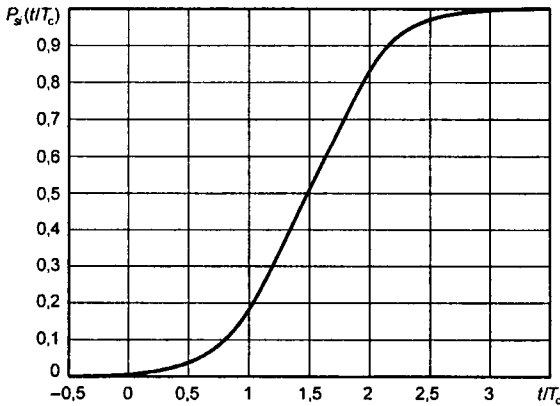


Рис. 6.17. Фазовая функция сигнала ГММС при $BT_c = 0,3$

вале $0 < T_s \leq 3,0$, длительность которого равна трем длительностям информационного символа. Вне этого интервала значения функции практически равны либо 0, либо 1.

В дальнейшем нам потребуется представление функции (6.34) на конечном интервале $k < t \leq k + 1$, на котором передается информационный символ с номером k . В соответствии с (6.34) сначала можно записать выражение

$$\begin{aligned} \varphi(t) &= \int_{-\infty}^t p(\tau) d\tau = \sum_{i=-\infty}^k x_i \int_{-\infty}^t q(\tau - i) d\tau = \\ &= \sum_{i=-\infty}^k x_i \psi(t - i), \quad k < t \leq k + 1, \end{aligned} \quad (6.36)$$

которое учитывает, что мы не рассматриваем символы со значениями индекса i , превышающими значение k . Однако это выражение можно представить в более простой форме, если учесть, что символы на прошлых интервалах оказывают разное влияние на значения функции $\varphi(t)$ на рассматриваемом интервале с номером k . Действительно, из рис. 6.17 следует, что все информационные символы со значениями индекса $i \leq k - 3$ добавляют в выражение (6.36) только $+1$ или -1 в соответствии со значениями символов x_i . Вклад двух предшествующих символов можно представить слагаемым $\sum_{i=k-2}^{k-1} x_i \psi(t - i)$. Тогда вместо (6.36) можно записать

$$\varphi(t) = \sum_{i=-\infty}^{k-3} x_i + \sum_{i=k-2}^{k-1} x_i \psi(t - i) + x_k \psi(t - i), \quad k < t \leq k + 1. \quad (6.37)$$

Значением первого слагаемого в правой части равенства здесь может оказаться любое целое число, положительное или отрицательное, включая нуль. Два последних слагаемых определяют изменение функции $\varphi(t)$ на интервале с номером k : последнее слагаемое — вклад текущего информационного символа на данном интервале, второе слагаемое — вклад двух предшествующих информационных символов в значения функции $\varphi(t)$ на текущем интервале.

Таким образом, сигнал на выходе устройства интегрирования обладает «памятью» — значения этого сигнала на интервале с номером k зависят от информационных символов на всех предшествующих интервалах.

Теперь возвращаемся к схеме модулятора (см. рис. 6.14), в соответствии с которой над сигналом $\varphi(t)$ одновременно выполняются следующие преобразования:

$$\cos\left[\frac{\pi}{2}\varphi(t)\right] = I(t), \quad \sin\left[\frac{\pi}{2}\varphi(t)\right] = Q(t). \quad (6.38)$$

В результате радиосигнал на выходе модулятора теперь принимает вид

$$s[t; u(t)] = \cos\left[\frac{\pi}{2}\varphi(t)\right] \cos(2\pi f_0 t) - \sin\left[\frac{\pi}{2}\varphi(t)\right] \sin(2\pi f_0 t) = \cos\left[2\pi f_0 t + \frac{\pi}{2}\varphi(t)\right]. \quad (6.39)$$

Здесь добавлен множитель $\pi/2$ для того, чтобы полное приращение фазы сигнала, обусловленное одним информационным символом, было равно $\pi/2$ (как это имеет место для ММС сигнала). Именно благодаря соответствующему свойству ММС сигнала рассматриваемый здесь сигнал также назван сигналом с минимальным сдвигом.

В согласии с (6.39) радиосигнал ГММС является сигналом с угловой (частотной или фазовой) модуляцией и имеет постоянное значение огибающей. Если рассматривать значения данного сигнала на всех интервалах вплоть до текущего интервала с номером k , то вся последовательность информационных символов $\{b_i, i = \dots, (k-1), k\}$ закодирована в значения функции $\varphi(t)$ на этих интервалах.

Формула (6.37) может быть использована для построения графиков фазовых траекторий сигнала ГММС при различных значениях произведения BT_s , которые можно устанавливать путем выбора соответствующего значения полосы пропускания B гауссовского фильтра при фиксированном значении длительности T_s информационного символа. В частности, при $B \rightarrow \infty$ полоса фильтра неограниченно возрастает, так что в пределе можно считать, что фильтр отсутствует и сглаживания фазовых траекторий не происходит. В результате приходим к сигналу с непрерывными фазовыми траекториями, которые в согласии с (6.37) на интервале $k < t \leq k+1$ представляются выражением

$$\varphi_\infty(t) = \sum_{i=-\infty}^{k-1} x_i + x_k \psi_\infty(t-k), \quad k < t \leq k+1, \quad (6.40)$$

где

$$\psi_\infty(t) = \begin{cases} 1, & \text{если } t > 1, \\ t, & \text{при } 0 \leq t \leq 1, \\ 0, & \text{если } t < 0. \end{cases} \quad (6.41)$$

Из (6.41) следует, что фазовая траектория в рассматриваемом случае является линейно ломанной: на каждом тактовом интервале фаза сигнала линейно изменяется ровно на $+90^\circ$ или -90° относительно фазы несущего колебания в зависимости от значения информационного символа, передаваемого на данном интервале. Поэтому фазовая решетка сигнала ГММС при $BT_s = \infty$ совпадает с фазовой решеткой сигнала ММС, представленной на рис. 6.9. Одна фазовая траектория сигнала выделена на рис. 6.9. Соответствующая ей фазовая траектория сигнала ГММС при $BT_s = 10$ и $BT_s = 0,3$ изображены на рис. 6.18, *a* и *b* соответственно. Сравнивая фазовые траектории, можно сделать вывод об идентичности их зависимости от информационных символов; однако фазовая траектория при $BT_s = 0,3$ является гладкой, что должно обеспечивать дополнительное уменьшение внеполосных излучений по сравнению с сигналом ММС.

Энергетический спектр сигнала ММС определяется соотношением (6.26). К сожалению, аналогичной явной формулы для спектральной плотности мощности сигнала ГММС не существует. Путем моделирования получены лишь оценки этого спектра, представляемые графиками при различных значениях параметра BT_s . Примеры таких графиков приведены на рис. 6.19.

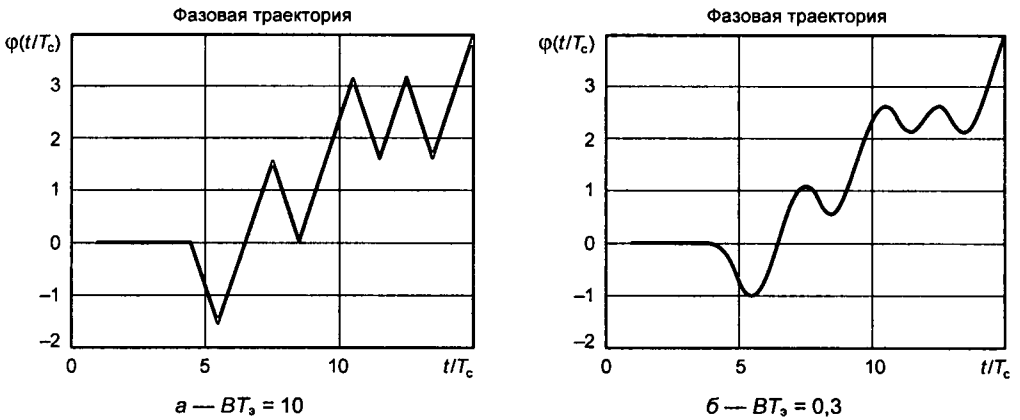


Рис. 6.18. Фазовые траектории сигнала ГММС

На рис. 6.19, *а* изображен спектр сигнала ГММС при $BT_s = 10,0$, который останется практически таким же при дальнейшем увеличении этого произведения. Поэтому график можно рассматривать как оценку спектра сигнала ММС, полученную путем имитационного моделирования; такую оценку можно сравнить с соответствующим графиком на рис. 6.13 построенным по аналитической формуле (6.26). Из рис. 6.19, *б* следует, что на уровне -40 дБ ширина спектра сигнала ГММС примерно в три раза уже, чем для сигнала ММС.

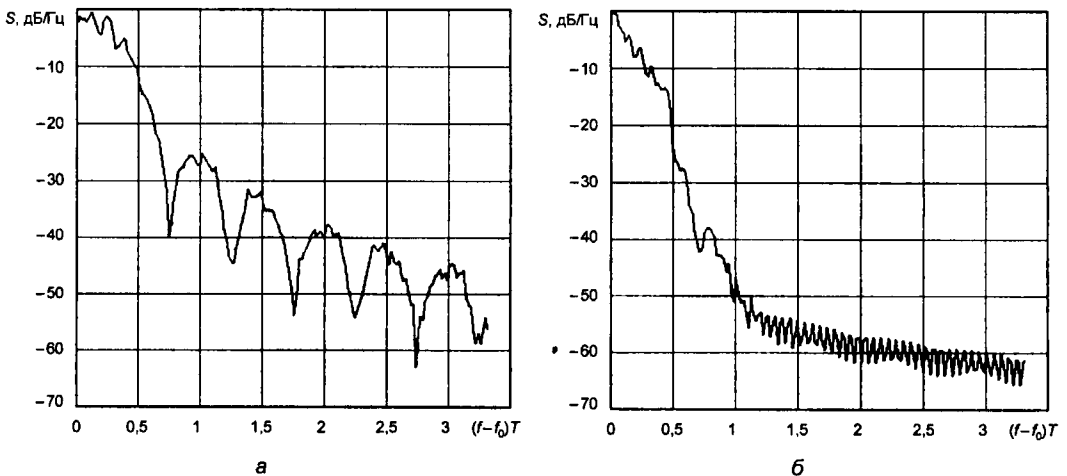


Рис. 6.19. Энергетический спектр сигнала ГММС

Таким образом, можно утверждать, что сглаживание фазовых траекторий сигналов с непрерывной фазой позволяет заметно повысить спектральную эффективность системы цифровой радиосвязи.

Вместе с тем приведенные данные указывают на возникновение значительных МСИ, вызванных использованием фильтрации закона изменения частоты при ГММС. Так, в соответствии с данными рис. 6.18 изменение фазы за время передачи данного символа существ-

венно зависит от символов, передававшихся ранее. Это означает, что можно ожидать существенного снижения помехоустойчивости сигналов ГММС по сравнению с ММС. Чтобы соответствующие энергетические потери были бы менее значительными, необходимо использовать метод демодуляции, оптимизированный с учетом наличия МСИ. Поскольку МСИ приводит к взаимозависимости сигналов, передаваемых на соседних тактовых интервалах, такая оптимизация (см. гл. 3), сводится к необходимости укрупнения ансамбля ЭС. Типовым способом решения данной задачи без недопустимого усложнения оборудования является использование алгоритма Витерби при демодуляции. Последний широко реализуется в ряде случаев для борьбы с искажениями сигналов, вызванных многолучевым распространением в канале связи (более подробно см. 9.4.2).

6.2.5. Подавление внеполосных излучений сигналов ФМ путем их фильтрации в передатчике

Использование фильтров, пропускающих основную полосу радиосигнала и подавляющих внеполосные составляющие его спектра, является наиболее очевидным путем решения интересующей нас задачи. Соответствующая фильтрация должна осуществляться на передающей стороне радиолинии.

Следует иметь в виду, что указанная процедура сопряжена с некоторыми важными особенностями.

Во-первых, указанная фильтрация практически всегда вызывает появление изменений амплитуды сигнала на выходе фильтра, так что в результате сигнал оказывается промодулированным не только по фазе, но и по амплитуде. Нетрудно убедиться, например, что если на вход полосового фильтра подать колебание его резонансной частоты со скачком фазы на 180° , то на выходе этот скачок сохранится, но в момент скачка амплитуда выходного сигнала будет равна нулю, так что, по крайней мере, первая производная выходного сигнала будет всегда существовать. Рассуждения, приводящие к такому выводу, иллюстрируются графически диаграммами рис. 6.20.

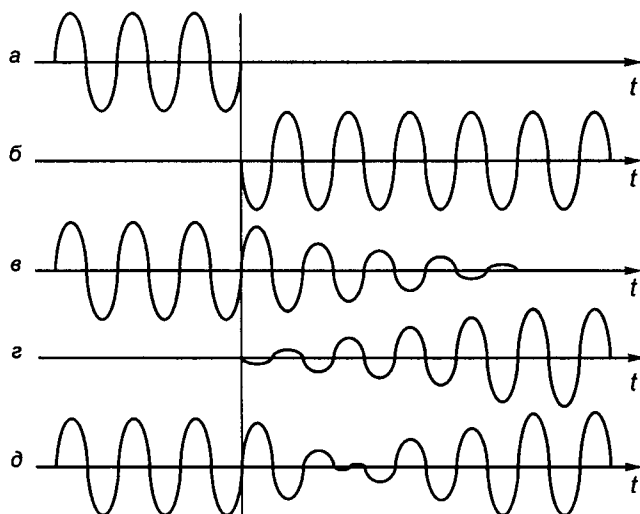


Рис. 6.20. К прохождению сигнала со скачком фазы через полосовой фильтр

Здесь входной сигнал фильтра со скачком фазы представлен как суперпозиция двух составляющих: «а» и «б»; отклики фильтра на каждую из них представлены на графиках «в» и «г» соответственно, а их сумма (выходной сигнал фильтра) — на графике «д». Когда отклик «в», спадая, станет равным по амплитуде нарастающему отклику «г», в выходном сигнале произойдет скачок фазы. Амплитуда выходного сигнала в этот момент времени будет равна нулю. Таким образом, на выходе фильтра амплитуда перестает быть постоянной и даже достигает в отдельные моменты времени нуля. Именно нулевая амплитуда в момент скачка фазы выходного сигнала фильтра приводит к появлению в данной точке производной сигнала и, как следствие, определяет эффект снижения внеполосных излучений за счет фильтрации.

Если выходной сигнал фильтра пропустить через ограничитель амплитуды, устраняющий возникшую амплитудную модуляцию, сигнал вновь станет таким же, как до фильтрации, а следовательно, восстановятся и все внеполосные составляющие. Таким образом, необходимо следить за тем, чтобы в передающем тракте после фильтрации использовались только линейные преобразования. В частности, и усилитель мощности должен работать в квазилинейном режиме, что сопряжено с существенным снижением выходной мощности и КПД передатчика. В этом заключается существенный недостаток фильтрации, как метода подавления внеполосных составляющих спектра сигнала.

Во-вторых, фильтрация всегда приводит к появлению межсимвольной интерференции в сигнале. Речь идет о том, что «хвосты» откликов фильтра на элементарные посылки, предшествующие данной, накладываются на нее и мешают ее приему. В результате может существенно увеличиться требуемое значение h_6 , т.е. снизиться энергетическая эффективность манипуляции.

Эти недостатки фильтровых методов подавления внеполосных составляющих спектра приводят к тому, что они используются лишь в тех случаях, когда эффект сглаживания скачков фазы оказывается недостаточным для достижения требуемого значения γ . Как отмечалось выше, ММС обеспечивает значение γ порядка единицы. В настоящем разделе будет показано, что рациональная фильтрация позволяет ориентировочно повысить эту величину в два раза без снижения энергетической эффективности.

Если появление амплитудной модуляции характерно для фильтровых методов при использовании фильтров с любыми характеристиками, то рациональный выбор этих характеристик позволяет практически полностью устранить вредное влияние МСИ. Чтобы пояснить как это происходит, рассмотрим бинарную фазовую манипуляцию, реализуемую в видеоспектре. Пусть в качестве переносчика используется периодическая последовательность δ функций Дирака с периодом следования T (рис. 6.21, а). Положим, что данный переносчик подвергается двоичной фазовой манипуляции (помножается на соответствующий СТС с символами $a_k = \pm 1$, где k — номер символа: $k = \dots -1, 0, 1, \dots$), в результате чего образуется последовательность, представленная на рис. 6.21, б.

Далее эта последовательность пропускается через идеальный ФНЧ (фильтр нижних частот с П-образной частотной и линейной фазовой характеристиками) с частотой среза $F_c = 0,5/T$. Фильтр относится к классу физически не реализуемых и практически может быть создан лишь фильтр, приближающийся к требуемому, причем, чем более точное приближение будет достигаться, тем большим окажется запаздывание отклика этого фильтра по отношению к воздействию. Исключая из рассмотрения несущественное для данного анализа запаздывание и принимая соответственно фазовый набег в ФНЧ равным нулю на всех частотах, найдем для импульсного отклика идеального ФНЧ:

$$H(t) = \frac{\sin(\pi t/T)}{\pi t/T}. \quad (6.42)$$

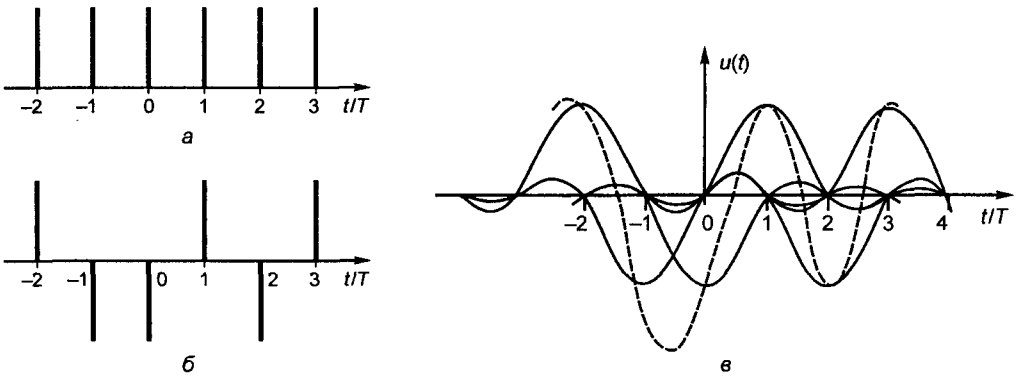


Рис. 6.21. Бинарная фазовая манипуляция в видеоспектре

Манипулированная последовательность на выходе ФНЧ примет вид

$$u(t) = \sum_k a_k H(t - kT) = \sum_k a_k \frac{\sin[\pi(t - kT)/T]}{\pi(t - kT)/T}. \tag{6.43}$$

Отдельные члены последнего выражения представлены на рис. 6.21, *в* сплошными линиями, а вся сумма — штриховой.

Из рис. 6.21 следует, что в тактовой точке $t = kT$, отличен от нуля только отклик фильтра на импульс a_k ; отклики на остальные импульсы манипулированного переносчика равны нулю. Поэтому при стробировании выходного сигнала фильтра точно в тактовых точках можно полностью разделить отдельные посылки и исключить межсимвольную интерференцию. При этом полоса частот, занимаемых выходным сигналом, строго равна нулю для частот, превышающих $F_c = 0,5/T$. Таким образом, внеполосные излучения оказываются полностью подавленными, что является прямым следствием использования при формировании сигнала на передающей стороне линии связи идеального ФНЧ. Использование такого фильтра, хотя и приводит к появлению протяженного импульсного отклика от каждой элементарной посылки, перекрывающего последующие посылки, но в тактовых точках все эти отклики равны нулю, так что межсимвольная интерференция отсутствует. В этом специальном смысле идеальный ФНЧ можно считать безинтерференционным фильтром.

Сформированный изложенным способом сигнал может применяться для передачи сообщений по каналу связи (рис. 6.22). До входа канала связи здесь представлена часть схемы формирующего сигнал $u(t)$ по описанному выше способу. В канале на данный сигнал накладывается белый гауссовский шум. Для подавления составляющих шума, частоты которых лежат вне спектра полезного сигнала, на приеме используется ФНЧ — полный аналог ФНЧ, установленного в тракте передачи. Сигнал с выхода ФНЧ поступает на схему стробирования, определяющую знак сигнала в каждой тактовой точке. Тем самым и завершается демодуляция принятого сигнала.

В блок-схеме на рис. 6.22 в тракт последовательно включены два фильтра — передачи и приема. Свойством безинтерференционности должен обладать эквивалентный фильтр, представляющий собой последовательное включение фильтров приема и передачи; обладают ли этим свойством каждый из указанных фильтров для наших целей несущественно. В рассмотренном варианте идеальных ФНЧ свойством безинтерференционности обладает как каждый фильтр в отдельности, так и их последовательное включение (подключение к идеальному ФНЧ последовательно любого числа таких же ФНЧ, образует тот же ФНЧ).

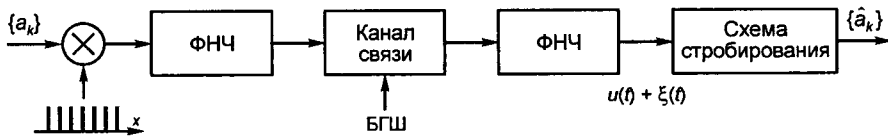


Рис. 6.22. Блок-схема рассматриваемого тракта передачи сообщений

Таким образом, в приведенной на рис. 6.22 линии связи на выходе стробирующего устройства присутствует полезный сигнал с уровнем -1 или $+1$ и шум с дисперсией $\sigma^2 = N_0 F_c = 0,5 N_0 / T$. Вероятность ошибки составляет соответственно

$$P_{\text{ош}} = \Phi(1/\sigma) = \Phi(\sqrt{2T/N_0}) = \Phi(\sqrt{2}h_g), \quad (6.44)$$

откуда следует, что помехоустойчивость описанного демодулятора совпадает с потенциальной помехоустойчивостью ФМ2 при когерентном приеме.

Что касается полосной эффективности описанной системы передачи, то для нее

$$\gamma_n = 2. \quad (6.45)$$

На основе описанного варианта модулированных видеосигналов легко сформировать и соответствующий радиосигнал. Так, сигнал ФМ2 может быть организован умножением $u(t)$ на несущую, а ФМ4 — формированием двух сигналов типа $u(t)$, манипулированных одним четными, а другой — нечетными символами сообщения и переноса этих видеосигналов на квадратурные несущие. Для получающихся радиосигналов (и ФМ2, и ФМ4) энергетическая эффективность определяется (6.44), т.е. совпадает с ФМ2, а полосная — для ФМ4 совпадает с (6.45), а для ФМ2 в два раза меньше. (Полоса частот занимаемая ФМ4 составляет $F = 0,5/T$, а ФМ — $F = 1/T$.)

Приведенные данные позволяют утверждать, что использование подавления внеполосных излучений за счет идеальной фильтрации позволяет повысить полосную эффективность в два раза по сравнению с ММС без снижения энергетической эффективности.

На практике реализуется лишь некоторое приближение к идеальному ФНЧ. В качестве типового приближения часто используются так называемые ФНЧ с косинус квадрат скруглением, частотная характеристика которых определяется соотношением

$$H(f) = \begin{cases} 1, & \text{при } |v| \leq 1-a; \\ \cos^2 \frac{\pi(v-1+a)}{4a}, & \text{при } 1-a < |v| < 1+a; \\ 0, & \text{вне этих интервалов,} \end{cases} \quad (6.46)$$

где $v = f/F_c$; F_c — частота среза.

Фильтры такого типа также относятся к физически не реализуемым, поскольку их частотная характеристика тождественно равна нулю в интервале частот. Поскольку в области среза спад характеристики происходит не скачком, как у идеальных ФНЧ, а плавно, приближение к ней осуществляется проще. Далее ФНЧ с АЧХ (6.46) являются безинтерференционными в определенном выше смысле. Можно показать, что этим свойством обладают все ФНЧ, частотная характеристика которых в области среза нечетна относительно частоты среза.

Следует иметь в виду, что поскольку характеристикой (6.46) должен обладать эквивалентный фильтр, образованный последовательным включением фильтров приема и передачи, то частотные характеристики последних должны иметь вид

$$H_1(f) = \sqrt{H(f)}. \quad (6.47)$$

6.2.6. Частотное разделение сигналов при использовании общего передающего тракта для многоканальной передачи

До сих пор мы рассматривали вопросы разделения сигналов и, в частности, частотного разделения в предположении, что общий тракт, используемый всеми каналами, является линейным. Это полностью оправдано, когда таким общим трактом является «эфир», остающийся линейным для сигналов достаточно большой мощности. Однако в других случаях, например, при использовании ретрансляторов для организации многоканальной передачи, это условие может и не выполняться. Передатчик ретранслятора при возрастании мощностей транслируемых сигналов рано или поздно переходит в нелинейный режим. Более того, усиление ретранслятора должно выбираться таким, чтобы обеспечить максимальную мощность выходных сигналов. Это может быть достигнуто путем приближения работы передатчика ретранслятора к режиму насыщения, до тех пор, пока потери, возникающие из-за нелинейных эффектов, не превысят выигрыша от соответствующего увеличения мощностей излучаемых ретранслятором сигналов. Эти соображения указывают на необходимость изучения влияния таких нелинейных эффектов.

Основным источником нелинейности ретрансляторов являются выходные каскады усилителей мощности (УМ). Именно анализ их нелинейности и составляет содержание настоящего раздела.

Из множества возможных нелинейных явлений выделим те, которые проявляются в первую очередь при переходе из линейного режима работы УМ в нелинейный. К ним относятся:

- нелинейность амплитудной характеристики (АХ) усилителя;
- зависимость сдвига фаз усиливаемого усилителем радиосигнала от амплитуды сигнала.

Так, АХ усилителя радиосигналов измеряется при подаче на его вход гармонического колебания и представляет собой зависимость амплитуды или мощности выходного сигнала УМ от того же параметра его входного колебания.

Понятие АХ применимо лишь к резонансным УМ, включающим в свой состав необходимые элементы фильтрации, так что при подаче на вход радиосигнала в некоторой полосе рабочих частот на выходе получаем искаженный радиосигнал в той же полосе. Таким образом, АХ определяет безынерционную нелинейность усилителя.

Идеальная АХ линейного усилителя представлена прямой 1 на рис. 6.23. Реальные характеристики отклоняются от идеальной: прежде всего, возникает загиб в области больших входных уровней (кривая 2), означающий переход к режиму насыщения. В качестве числовой характеристики отклонения используют уровень компрессии $q = P_0/P$ дБ. Говорят, например, что при выходной мощности УМ, равной P , уровень компрессии составляет q дБ.

Уровень компрессии далеко не всегда в достаточной мере характеризует нелинейность АХ. Из многочисленных случаев, когда эта характеристика оказывается недостаточной, особенно существен часто встречающийся вариант, когда АХ имеет выраженную нижнюю пятку (кривая 3). Такого рода характеристики типичны, например, для твердотельных усилителей, работающих в режимах между классами С и В. В связи с этим используют более сложные числовые характеристики нелинейности АХ, основанные на

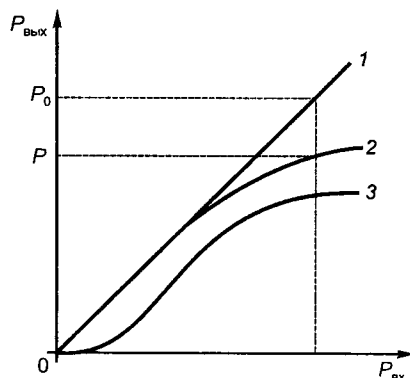


Рис. 6.23. Амплитудная характеристика усилителя

непосредственном измерении отрицательных эффектов, к которым такая нелинейность приводит.

Эти эффекты проявляются, когда одновременно усиливаются не один, а несколько гармонических колебаний различных частот. В выходном сигнале появляются так называемые комбинационные составляющие, частоты которых не были представлены во входном сигнале, а получаются как линейные комбинации частот входных гармонических колебаний. Остановимся на данном вопросе подробнее.

Пусть на вход УМ с нелинейной АХ подадут сумму двух гармонических колебаний частот $f_1 = f_0 + F_1$ и $f_2 = f_0 + F_2$, где f_0 — средняя частота усиливаемой полосы рабочих частот, а величины F определяют сдвиг рассматриваемых составляющих, относительно данной частоты. Вычисляя огибающую суммы поданных составляющих, преобразуя ее в соответствии с АХ усилителя и вычисляя спектр его выходного сигнала, можно убедиться, что в общем случае наблюдается следующее:

- в выходном сигнале кроме гармонических составляющих частот f_1 и f_2 появляются также составляющие с частотами $f_k = f_0 + m_1 F_1 + m_2 F_2$, где m_j — любые целые числа положительные или отрицательные, причем $|m_1 + m_2| = 1$; эти составляющие и именуется комбинационными;
- комбинационные составляющие отбирают часть мощности выходного сигнала, вследствие чего уменьшается мощность полезного выходного сигнала;
- при усилении модулированных радиосигналов, возникающие комбинационные составляющие, в общем случае перекрываются по спектру с полезными колебаниями и тем самым создают помехи их приему;
- при прохождении через УМ отношение уровней усиливаемых полезных сигналов изменяется; причем часто это изменение может происходить в пользу более сильного сигнала; тогда это явление именуют *подавлением сильным сигналом слабого*.

В более общем случае, когда входной сигнал содержит более двух гармонических компонент, частоты возникающих комбинационных составляющих определяются суммой

$$f_k = \sum_j m_j f_j \quad \text{при условии, что} \quad \sum_j m_j = 1. \quad (6.48)$$

Числовой характеристикой той или иной комбинационной составляющей является ее порядок, по определению равный $n_n = \sum_j |m_j|$. В соответствии с (6.48) порядок может быть

только нечетным целым положительным числом не меньше 3. (Нечетность порядка следует из того факта, что для выполнения (6.48) необходимо, чтобы среди всех m_j было бы нечетное число нечетных чисел.) С ростом порядка уровень комбинационных компонент снижается, хотя это снижение может и не быть монотонным. Тем не менее максимальными бывают, как правило, комбинации третьего порядка.

Имея в виду количественную оценку перечисленных выше отрицательных явлений, связанных с комбинационными искажениями, остановимся на некоторых типовых примерах.

Первый из них относится к случаю, когда число усиливаемых радиосигналов невелико, и УМ работает в режиме близком к насыщению. Приемлемой моделью для такого случая служит входной двухчастотный сигнал и АХ, соответствующая идеальному ограничителю амплитуд, представленная на рис. 6.24.

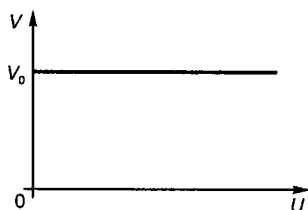


Рис. 6.24. Амплитудная характеристика УМ:
U, V — огибающие входного и выходного сигналов

Поскольку УМ предполагается безынерционным, то он без искажений воспроизводит текущую фазу $Q(t)$ входного колебания. Поэтому, если входной сигнал представлен в виде $x(t) = U(t) \cos Q(t)$, то выходной сигнал УМ имеет вид

$$y(t) = V(t) \cos Q(t) = V_0 \cos Q(t). \quad (6.49)$$

Итак, пусть на вход УМ с АХ (см. рис. 6.24) воздействует сумма (биения) двух гармонических колебаний с амплитудами U_1 и U_2 и частотами f_1 и f_2 . Рассмотрим вначале случай, когда отношение амплитуд $q = U_2/U_1$ достаточно мало. Тогда, как это следует из теории биений, суммарный сигнал представляет собой амплитудно и частотно модулированное по синусоидальному закону колебание с мгновенной частотой

$$f(t) = \frac{1}{2\pi} \frac{d}{dt} Q(t) = f_1 + q(f_2 - f_1) \cos[2\pi(f_2 - f_1)t]. \quad (6.50)$$

В данном случае УМ устраняет изменения амплитуды, так что его выходной сигнал представляет собой колебание постоянной амплитуды V_0 , модулированное по частоте синусоидальным напряжением частоты $f_2 - f_1$ с малым индексом модуляции q . Спектр такого колебания состоит (с точностью до величин второго порядка малости) из несущей частоты f_1 , с амплитудой V_0 и двух боковых с амплитудой каждой $V_0 = 0,5q V_0$ и частотами: для верхней боковой — $f_1 + (f_2 - f_1) = f_2$ и нижней — $f_1 - (f_2 - f_1) = f_0 + 2f_1 - f_2 = f_k$. Несущая представляет собой усиленное первое колебание, верхняя боковая — усиленное второе, а нижняя боковая — комбинационную составляющую третьего порядка. Частота последней зеркальна относительно f_1 частоте f_2 . Приведенные данные по уровням составляющих спектра показывают, что в усилителе, работающем в режиме насыщения (идеального ограничителя амплитуд), наблюдается подавление слабого сигнала сильным. Отношение амплитуд данных сигналов на выходе усилителя в два раза отличается от их отношения на его входе. Таким образом, коэффициент подавления составляет 6 дБ. Кроме того, в усилителе возникает комбинационная составляющая третьего порядка, уровень которой равен выходному уровню слабого сигнала.

Описанные явления полностью сохраняются, если вместо двух гармонических колебаний на вход усилителя подать два модулированных сигнала с существенно различными уровнями или один сигнал большого уровня и несколько слабых сигналов, таких, что их суммарный уровень значительно меньше большого сигнала. В последнем случае сумма всех слабых сигналов может рассматриваться как один сложный слабый сигнал.

Если в общем случае и сильный, и слабый сигналы имеют амплитудную модуляцию, то q изменяется во времени по закону, учитывающему составляющие данной модуляции обоих сигналов. При этом (6.50) остается справедливым, если только в любой момент времени q достаточно мало. Это означает, что в дополнение к отмеченным явлениям подавления слабого сигнала и возникновения комбинационной составляющей происходит еще и следующее: амплитудная модуляция сильного сигнала устраняется, и она переходит на слабый сигнал с инвертированием закона модуляции.

Рассмотрим ситуацию, когда входной сигнал состоит из большого числа независимых радиосигналов различных каналов. Тогда распределение вероятностей $x(t)$ можно считать нормальным, так что огибающая $U(t)$ распределена по закону Релея:

$$p(U) = \frac{U}{P_{\text{вх}}} \exp(-U/2P_{\text{вх}}), \quad (6.51)$$

где $P_{\text{вх}}$ — мощность входного сигнала.

Пусть АХ имеет вид $V = L(U)$, где L — функция, описывающая эту характеристику. Сигнал y состоит из усиленного в k раз сигнала x и комбинационных искажений, сумму которых обозначим как z :

$$y(t) = kx(t) + z(t). \quad (6.52)$$

Случайные процессы x и z некоррелированы (в противном случае из z можно было бы еще выделить составляющие усиливаемого сигнала). Величина k зависит от уровня входного сигнала и вида АХ и может быть определена из следующих соотношений.

Коэффициент корреляции x и y по определению этих величин составляет

$$r = \overline{x(t)y(t)} = \overline{V(t)U(t)\cos^2 Q(t)} = \frac{1}{2} \overline{U(t)L[U(t)]}, \quad (6.53)$$

причем здесь и ниже усреднение должно производиться по распределению (6.51).

Таким образом, вычисляется величина r . В соответствии с (6.52) имеем $\overline{x(t)y(t)} = r = \overline{kx^2(t) + x(t)z(t)} = kP_{\text{вх}}$, так что коэффициент усиления определяется соотношением

$$k = r/P_{\text{вх}}. \quad (6.54)$$

Теперь может быть вычислена мощность полезного сигнала на выходе:

$$P_{\text{вых}} = k^2 P_{\text{вх}} = r^2/P_{\text{вх}}. \quad (6.55)$$

Выразим мощность всего выходного сигнала через P_{Σ} :

$$P_{\Sigma} = \overline{V^2(t)\cos^2 Q(t)} = \frac{1}{2} \overline{D^2[U(t)]}. \quad (6.56)$$

Обозначая мощность комбинационных помех на выходе через $P_{\text{кп}}$, найдем

$$P_{\text{кп}} = P_{\Sigma} - P_{\text{вых}}. \quad (6.57)$$

Уровень комбинационных помех оценивается по отношению к общей выходной ВЧ-мощности усилителя. Соответствующий коэффициент составляет

$$\beta^2 = \frac{P_{\text{кп}}}{P_{\Sigma}} = 1 - \frac{P_{\text{вых}}}{P_{\Sigma}}. \quad (6.58)$$

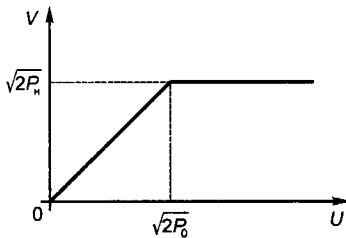


Рис. 6.25. Идеализированная АХ усилителя:

P_0 — входная мощность, соответствующая режиму насыщения;
 P_n — выходная мощность в режиме насыщения

Если не аналитически, то по крайней мере численно соответствующие расчеты могут быть всегда выполнены. В качестве примера рассмотрим случай, когда АХ может быть аппроксимирована идеализированной зависимостью, представленной на рис. 6.25.

После вычислений по приведенным выше формулам, найдем

$$\frac{r}{\sqrt{P_n}} = \frac{1 - \exp(-\alpha^2)}{\alpha} + \sqrt{\pi} \Phi(\sqrt{2}\alpha); \quad (6.59)$$

$$w_y^2 = \frac{P_{\Sigma}}{P_n} = \frac{1 - \exp(-\alpha^2)}{\alpha^2}, \quad (6.60)$$

где $\alpha^2 = P_0/P_{\text{вх}}$ — относительный уровень возбуждения усилителя; $\Phi(x)$ — функция ошибок (см. (1.40)). Результаты вычислений по этим формулам представлены в табл. 6.1.

Таблица 6.1. Уровни сигнала и комбинационных помех на выходе усилителя

α	0	0,35	0,53	0,7	1,06	1,41	2,12	$\rightarrow \infty$
w_y	1	0,96	0,86	0,8	0,6	0,42	0,22	$1/\alpha^2$
w_x	0,76	0,8	0,72	0,72	0,56	0,4	0,22	$1/\alpha^2$
β , дБ	-6,8	-7,45	-7,96	-10	-11,55	-14,56	-25,6	$10 \lg[(\exp - \alpha^2)/\alpha^2]$

Здесь представлены простые асимптотические формулы, которые относятся к случаям работы усилителя в квазилинейном режиме.

Интересно сопоставить данные табл. 6.1 с известными практическими результатами. АХ в некоторой степени, подобные представленной на рис. 6.25, имеют усилители на ЛБВ. Обычно в таких усилителях, для достижения значения $\beta = -20$ дБ приходится снижать выходную ВЧ-мощность примерно на 6 дБ ($w_y = 0,25$), что достаточно хорошо соответствует данным таблицы.

Оценим комбинационные искажения, возникающие вследствие другого нелинейного эффекта — зависимости сдвига фазы в усилителе от амплитуды усиливаемого радиосигнала. Соответствующая зависимость описывается амплитудно-фазовой характеристикой (АФХ). Для примера на рис. 6.26 представлены типовые АФХ для усилителей на ЛБВ (штриховая линия) и твердотельных СВЧ-усилителей (сплошная). По оси абсцисс отложен уровень входного сигнала УМ в дБ относительно уровня, переводящего усилитель в режим насыщения; для твердотельного усилителя насыщением считается точка на АХ этого усилителя, соответствующая уровню компрессии 2 дБ. Отложенный на этих рисунках сдвиг фазы $\Delta\varphi$ на самом деле полностью не определен. Дело в том, что в СВЧ-усилителях полный сдвиг фазы может во много раз превосходить 2π , так что измерение этой величины отнюдь не является тривиальным. Зато изменение сдвигов фазы в усилителе от тех или иных причин часто оказывается гораздо меньше 2π , так что их измерение оказывается более простым. Это в полной мере относится и к зависимости сдвига фазы от уровня входного сигнала. На рис. 6.26 представлено изменение сдвига фазы, причем за нуль принят сдвиг фазы при уровне сигнала минус 20 дБ.

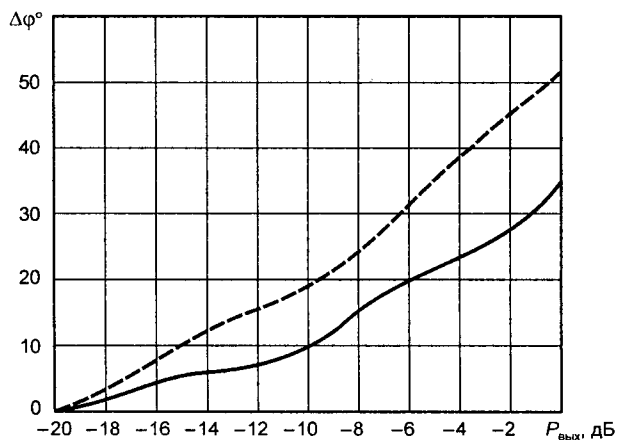


Рис. 6.26. Типовые АФХ усилителей мощности

Чем больше крутизна АФХ, тем в большей степени проявляется рассматриваемый нелинейный эффект. Поэтому в качестве его простой числовой характеристики часто рассматривается максимальное значение модуля крутизны, называемое коэффициентом амплитудно-фазового перехода $\alpha_{\text{АФФ}}, \text{ }^\circ/\text{дБ}$.

Встречающиеся на практике значения $\Delta\varphi^\circ$ оказываются небольшими по сравнению с 90° . С учетом данного обстоятельства для выходного сигнала усилителя можно записать в первом приближении:

$$z(t) = V(t) \cos [Q(t) + \Delta\varphi(t)] = V(t) \cos Q(t) + V(t) \frac{\Delta\varphi}{57,3} \sin Q(t), \quad (6.61)$$

$\Delta\varphi$ измеряется в градусах.

Мощность комбинационных искажений может быть определена из (6.61), если задаться амплитудой входного сигнала $U(t)$, коэффициентом усиления усилителя $k = V/U$ и АФХ, т.е. зависимостью $\Delta\varphi$ от U . В качестве примера выполним соответствующий расчет для случая, когда на вход усилителя воздействует групповой многоканальный сигнал с частотным делением, который может быть аппроксимирован нормальным шумовым случайным процессом. Что касается вида АФХ, то в соответствии с данными рис. 6.26 она в системе координат «уровень входного сигнала в дБ – фазовый сдвиг в градусах» может быть грубо аппроксимирована прямой. Именно:

$$\Delta\varphi = 20\alpha_{\text{АФХ}} \lg(U/U_0). \quad (6.62)$$

Здесь коэффициент 20 обеспечивает перевод уровня входного сигнала в дБ, а U_0 определяет постоянный сдвиг АФХ вдоль оси абсцисс. Как уже отмечалось выше, последний не влияет на комбинационные искажения и может быть выбран произвольно исходя из удобства расчета. Выберем U_0 таким образом, чтобы второй член выражения (6.61) не содержал полезного сигнала, а включал только продукты нелинейных искажений. Для этого U_0 следует выбрать из условия некоррелированности огибающих первого и второго членов (6.61),

что эквивалентно требованию $(U) \left(U \lg \frac{U}{U_0} \right) = 0$. Используя распределение (6.51) и таблич-

ные интегралы, найдем, что из этого условия следует $U_0 = 1,53 \sqrt{P_{\text{вх}}}$. При таком выборе суммарная мощность комбинационных искажений $P_{\text{кп}}$ определяется мощностью составляющей, отображаемой вторым членом выражения (6.61), а полезного выходного сигнала $P_{\text{вых}}$ — первого его члена. С применением того же распределения и табличных интегралов найдем

$$\frac{P_{\text{кп}}}{P_{\text{вых}}} [\text{дБ}] = 20 \lg \alpha_{\text{АФХ}} - 20,5. \quad (6.63)$$

В последнем соотношении следует в качестве $\alpha_{\text{АФХ}}$ подставлять среднее значение крутизны АФХ на интервале изменений огибающей входного группового сигнала. Для случая шумового воздействия основной вклад в комбинационные искажения дают изменение уровня входного сигнала от $0,25 P_{\text{вх}}$ до $4P_{\text{вх}}$, т.е. в диапазоне 12 дБ. Если воспользоваться рис. 6.26, то можно убедиться, что среднее значение $\alpha_{\text{АФХ}}$ в таком диапазоне составляет: для УМ на ЛБВ (2–3) °/дБ, а для твердотельных УМ (1,5–2) °/дБ, причем большие значения соответствуют выбору рабочей точки ближе к режиму насыщения, т.е. лучшему использованию мощности усилителя. С учетом этих данных на основании выражения (6.63) можно сделать вывод о том, что комбинационные составляющие, вызванные зависимостью фазового набега в усилителе от амплитуды, имеют тот же порядок, что и искажения, вызванные нелинейностью АХ.

Влияние возникающих из-за нелинейных явлений в УМ комбинационных составляющих на трансляцию группового многоканального сигнала сводится к двум эффектам:

- отбору части выходной ВЧ мощности на комбинационные составляющие за счет соответствующего уменьшения мощности полезного сигнала;
- уменьшению отношения полезный сигнал/помехи в связи с добавлением к шумам еще и комбинационных помех.

Выше ограничивались определением суммарной мощности всех комбинационных составляющих. Данных достаточно, чтобы охарактеризовать количественно первый из ука-

занных эффектов, но не достаточно для оценки второго. В последнем случае необходимо знать еще энергетический спектр комбинационных составляющих. Грубо он может быть определен на основании следующих рассуждений. Как уже отмечалось, наибольшие уровни имеют комбинации третьего порядка, частоты которых образуются по закону

$$f^* = f_i + f_j - f_k, \quad (6.64)$$

где i, j, k — номера некоторых частотных составляющих из спектра группового сигнала (ГС). Здесь в качестве f_i и f_j может выступать одна и та же частота. Легко понять, что полоса частот, занятых комбинациями третьего порядка, в три раза превосходит полосу ГС Δf (рис. 6.27, штриховой линией обозначен энергетический спектр ГС). Число возможных различных наборов частот, дающих комбинацию данной частоты, тем больше, чем ближе f^* к центру полосы ГС. Для f^* , отстоящих от этого центра на $1,5 \Delta f$, имеется всего один набор дающих такую комбинацию частот. Соответственно энергетический спектр комбинационных составляющих и приобретает форму, показанную сплошной кривой на рис. 6.27. Соответственно, энергетическая спектральная плотность комбинационных помех имеет максимум в центре полосы ГС, составляющий ориентировочно

$$S_m = \frac{2P_{\text{кп}}}{3\Delta f}, \quad (6.65)$$

и спадающий в два раза к краю этой полосы.

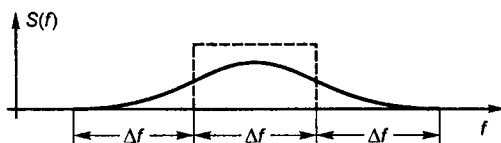


Рис. 6.27. Вид энергетического спектра комбинационных искажений

Следует иметь в виду, что приведенные данные характеризуют спектр нелинейных искажений лишь в среднем. Если в спектре ГС имеются гармонические составляющие (чаще ими оказываются несущие частоты некоторых сигналов), то и в спектре комбинаций появляются гармонические составляющие. Это приводит к тому, что отдельные, весьма узкие участки спектра оказываются пораженными помехами в значительно большей степени, чем в среднем. Поэтому здесь должны использоваться специальные способы устранения гармонических составляющих в спектре сигналов. К ним относятся соответствующие способы манипуляции и преобразования манипулирующего сообщения.

6.3. Системы с временным разделением

6.3.1. Общие сведения

Способы организации временного разделения можно классифицировать, прежде всего, в зависимости от того, какой именно ресурс распределяется между отдельными каналами. В качестве такого ресурса может выступать либо общий видео-, либо общий радиоканал. В первом случае будем говорить о временном видеоделении (видеоуплотнении), а во втором — о временном радиоделении (радиоуплотнении)

Полагаем, что передаче подлежит m сообщений, представляющих собой двоичные цифровые сигналы с тактовыми частотами F_j ($j = 1, 2, \dots, m$). При видеоуплотнении все сообщения поступают на вход одной радиостанции и должны быть ею переданы совместно. На передающем конце радиолинии они объединяются по времени в единый групповой цифровой видеосигнал, который модулирует общую несущую и передается по радиолинии. На приемной стороне после демодуляции и выделения группового сигнала осуществляется его разделение на отдельные сообщения.

В отличие от данного варианта при радиоуплотнении отдельные сообщения передаются различными радиостанциями. На каждой станции подлежащее передаче сообщение пакетизируется, т.е. преобразуется в последовательность блоков (пакетов) символов, разделенных паузами. Пакеты следуют периодически и каждый символ сообщения включается в некоторый пакет и при том только один раз. Далее пакетированное сообщение манипулирует несущую, причем в паузах (между пакетами) сигнал данной радиостанцией не передается. В результате формируется пакетный радиосигнал, который и излучается в эфир.

Другие станции тем же способом на базе своих сообщений формируют собственные радиосигналы на той же (по номиналу) несущей частоте. Передача этих сигналов должна осуществляться таким образом, чтобы в месте приема пакеты различных станций не перекрывались во времени. Если прием в данной системе осуществляется в нескольких точках, то это требование должно выполняться для любой из них. При радиоуплотнении реализация отмеченного требования обеспечивается соответствующей системой синхронизации (системой синхронизации пакетов — ССП).

Технические решения, используемые при временном видео- и радиоуплотнении, имеют много общего. Для определенности ниже изложим их применительно к видеоуплотнению, после чего отметим некоторые особенности, относящиеся к радиоуплотнению.

Начнем с классификации подлежащих уплотнению сообщений. Назовем их *синхронными* в случаях, когда тактовые частоты всех сообщений строго кратны некоторой частоте F . В противном случае ансамбль сообщений будем именовать *асинхронным*. Соответственно, временное уплотнение, требующее, чтобы сообщения были синхронными, называют синхронным, а допускающее их возможную асинхронность — асинхронным.

Общая идея асинхронного уплотнения сводится к тому, что полные сведения о каждом из объединяемых сообщений передаются с помощью вспомогательных двоичных последовательностей — переносчиков, причем тактовые частоты последних отличаются от тактовых частот сообщений и формируются так, чтобы переносчики образовывали синхронный ансамбль. Сигналы этого ансамбля, после их синхронного объединения образуют ГС. Если, как мы всегда будем полагать, ГС является двоичным, то его тактовая частота не может быть меньше суммы тактовых частот сообщений.

Процедура преобразования сообщения в последовательности — переносчики (в общем случае для трансляции одного сообщения может использоваться несколько таких переносчиков), представляет собой специфическое кодирование, называемое *асинхронным вводом*. Обратная процедура, выполняемая при разделении, называется *асинхронным выводом*.

Принцип осуществления синхронного уплотнения проиллюстрируем на простейшем примере, когда тактовые частоты всех уплотняемых сообщений равны, т.е. для всех j $F_j = F$. Уплотнение может быть осуществлено простым коммутатором, способным подключать одну клемму (назовем ее *групповой*) к любой из $(m + 1)$ клемм, которые будем именовать *индивидуальными*. Сообщения подаются на индивидуальные клеммы, и коммутатор поочередно и периодически с частотой F подключает каждую из них к групповой клемме. Не занятая сообщениями $(m + 1)$ -я индивидуальная клемма используется для передачи маркерной последовательности, обозначающей начало очередного цикла опросов уплотняемых сообще-

ний. Маркерная последовательность, будучи заранее известной, позволяет при разделении идентифицировать каналы по их номерам.

Последовательность, образуемая в результате описанной процедуры на групповой клемме, и есть ГС. Его тактовая частота составляет $F_r = (m + 1)F$. Если принять, что групповая клемма подключена к каналу связи, то можно сказать, что этот канал предоставляется для трансляции каждого из сообщений на короткие временные интервалы длительностью $\tau = 1/F_r$ ровно F раз в секунду, так что при передаче за время τ одного двоичного символа, все символы сообщения могут быть переданы по одному разу.

Описанный вариант синхронного уплотнения является простейшим. В более общем случае тактовые частоты сообщений (включая маркерную последовательность) не равны, но синхронны, т.е. представимы в виде $F_j = b_j F$, где b_j — целые числа. При этом уплотнение может осуществляться коммутатором, аналогичным описанному, работающим с периодом (циклом) $T = 1/F$ и предоставляющим за цикл канал связи j -му сообщению ровно b_j раз. Тактовая частота ГС составляет здесь $F_r = \sum_j F_j$.

Что произойдет, если на некоторый вход (например, i -й) описанного коммутатора подать несинхронное сообщение с тактовой частотой F_i^* , несколько отличной от синхронного значения F_i ? В этом случае число предоставлений для данного сообщения канала связи будет отличаться от числа символов в нем за то же время, так что либо некоторые символы будут пропущены (при $F_i^* - F_i > 0$), либо будут переданы лишние символы (при $F_i^* - F_i < 0$). Соответствующие искажения при передаче цифровых сообщений именуют *вставками*, отрицательными в первом случае и положительными во втором. Вставки представляют собой весьма опасные искажения, поскольку каждая вставка приводит к срыву нормального приема целой группы следующих за ней символов. Дело в том, что, как правило, символы сообщения разбиваются на отдельные комбинации («слова») и тот или иной смысл приписывается не каждому отдельному символу, а целой комбинации. Примером могут служить байты для передачи цифро-буквенного текста в компьютерной технике. Каждая вставка приводит к возникновению ошибки в определении на приемной стороне границ «слов» и, следовательно, фактически к полному нарушению приема сообщения до тех пор, пока тем или иным способом правильное положение границ не будет восстановлено.

По этой причине в радиолиниях задают весьма жесткие требования по вероятности появления вставок: допустимый временной интервал между последовательно возникающими вставками измеряется часами, а на магистральных линиях — сутками. В связи с этим и приходится использовать асинхронные методы временного уплотнения, исключаящие появление вставок.

В качестве маркерного сигнала в системах с временным делением можно использовать любую синхронную последовательность с частотой F . Требуется лишь, чтобы ее вид был заранее известен на приемном конце линии связи. Однако используют периодическую маркерную последовательность, представляющую собой повторение с периодом T некоторой синхрогруппы элементарных символов. Соответственно отрезок группового сигнала между двумя такими синхрогруппами (включающий и саму синхрогруппу) называют *циклом*. Последний включает $N_{\text{ц}} = TF_r$ двоичных символов, содержащих информацию о сообщениях и синхрогруппе. Принадлежность каждого двоичного символа цикла тому или иному сообщению (каналу) или синхрогруппе определяет структуру цикла.

Зная вид синхрогруппы, можно обнаружить ее в принятом групповом сигнале. Эту задачу решает система цикловой синхронизации (СЦС), входящая в аппаратуру временного разделения. По известной структуре цикла можно выделить символы цикла, относящиеся к любому требуемому сообщению (каналу).

6.3.2. Цикловая синхронизация и скремблирование

Задача СЦС в соответствии с вышеизложенным заключается в определении временных положений синхрогрупп в групповом сигнале. Препятствуют этому как ошибочная демодуляция отдельных символов на приеме, вызванная неизбежными мешающими воздействиями, так и ложный набор синхрогруппы символами транслируемых сообщений.

Первый из указанных механизмов очевиден: сбой некоторых символов, входящих в синхрогруппу, делает невозможным ее опознание. Суть второго механизма требует определенных пояснений. Поскольку синхрогруппы содержат только служебную информацию, их процент в общем групповом потоке должен быть по возможности минимальным. В связи с этим используют короткие синхрогруппы, состоящие всего из нескольких или одного двоичного символа. Такие короткие комбинации могут с достаточно высокой вероятностью встречаться в информационной части цикла, даже когда уплотняемые сообщения являются чисто случайными. Основой разделения истинных и ложных синхрогрупп становится проверка периодичности их повторения: если истинные синхрогруппы повторяются в каждом цикле на одних и тех же позициях, то ложные, как правило, этим свойством не обладают.

Дело усложняется, когда сообщение носит не случайный характер, а содержит длительные отрезки с регулярной и, в частности, периодической структурой. Это может иметь место, например, при отсутствии информации, когда непрерывно передается единица или ноль, при многократной передаче сигналов вызова и т.п. В таких ситуациях и ложные синхрогруппы могут следовать периодически, так что нормальная работа СЦС станет невозможной.

В связи с этим в системах с временным делением на передающей стороне радиолинии групповой сигнал подвергается специальному безызыточному кодированию, направленному на разрушение регулярных связей в последовательности символов. Такое преобразование именуется *скремблированием*. Отметим, что в современных радиолиниях скремблирование используется почти всегда, независимо от методов уплотнения. Оно оказывается полезным по ряду причин и, в частности, разрушает регулярные структуры, которые могут появиться в групповом сигнале, приводит к нормализации числа знакоперемев в нем (изменений его параметров в тактовых точках), даже если символы транслируемого сообщения в этих точках не изменяются. Последнее необходимо для нормальной работы систем тактовой синхронизации приемников (СТС), т.е. устройств, определяющих временное положение тактовых точек принимаемого сигнала.

Наиболее эффективный вариант скремблирования связан с формированием специальной цифровой псевдослучайной последовательности (ПСП), аналогичной тем, которые (см. 6.1) используются при формировании ФМ ШПС. Для скремблирования формируют ПСП, совпадающую по такту с ГС и с периодом составляющим в простейшем случае один, а в более общем варианте — несколько циклов данного сигнала. Сформированная таким способом ПСП складывается по $\text{mod } 2$ с информационными символами ГС (символы, входящие в синхрогруппы, преобразованию не подвергаются). Образующийся в результате сигнал, именуемый скремблированным, передается по каналу связи.

На приемной стороне осуществляется обратное преобразование — дескремблирование. После того как с помощью СЦС будут обнаружены временные положения синхрогрупп принятого группового сигнала, в приемнике формируется та же ПСП, что и на передаче, и с ее помощью вновь осуществляется скремблирование принятого сигнала. При правильной фазировке ПСП повторное скремблирование означает дескремблирование, поскольку двукратное добавление по $\text{mod } 2$ одной и той же последовательности эквивалентно добавлению ко всем символам группового сигнала нуля, т.е. его неискаженному воспроизведению.

В ситуации, когда период ПСП совпадает с циклом, правильная фазировка ПСП обеспечивается работой СЦС. В более общем случае требуется еще определить, на каком именно цикле начинается период ПСП. Эта задача может трактоваться иным образом: пусть в ГС вводится еще и сверхцикл, объединяющий некоторое фиксированное число следующих подряд циклов. Причинами организации сверхциклов могут быть не только скремблирование, но и организация асинхронного ввода, объединение сигналов с резко различающимися тактовыми частотами и т.п. При этом возникает задача осуществления не только цикловой, но и сверхцикловой синхронизации, которую также решает СЦС. Возможность такого решения обеспечивается введением на передающем конце измененных синхрогрупп, отмечающих начало сверхцикла. Например, если «рядовые» циклы отмечаются некоторой синхрогруппой, то первый цикл сверхцикла отмечается ее инверсией.

Важным свойством описанного способа скремблирования является то, что он не размножает ошибки, возникающие в канале, связи: если в приемнике некоторый символ ГС был демодулирован ошибочно, то после дескремблирования этот символ и только он будет восстановлен неверно. С другой стороны, из-за необходимости синхронизации по периоду ПСП предлагаемый метод скремблирования сравнительно сложный. При временном разделении это обстоятельство не проявляется, поскольку соответствующая система синхронизации объединяется с СЦС. В общем же случае, когда она должна специально создаваться для обеспечения скремблирования, усложнение может представляться чрезмерным. Тогда используются другие варианты скремблирования, хотя и приводящие к размножению ошибок, но не требующие дополнительной цикловой синхронизации.

Такие системы скремблирования основаны на использовании цифровых преобразователей (ЦП), представляющих собой регистр сдвига с несколькими отводами (рис. 6.28), причем в некоторых отводах устанавливаются инверторы («минус» в кружке). Последовательности, поступающие с отводов, суммируются по mod 2. Выход сумматора служит выходом преобразователя. Ясно, что если в ЦП используется регистр достаточно большой длины и с достаточно большим числом отводов, то соответствующее преобразование разрушает регулярные связи между символами последовательности, приближая ее к случайному телеграфному сигналу.

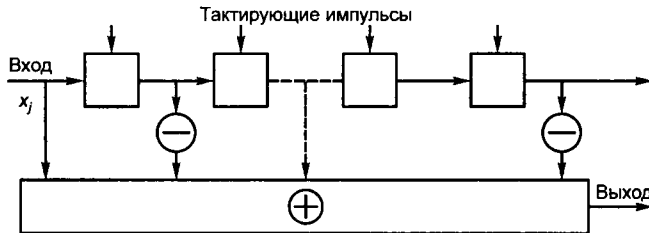


Рис. 6.28. Цифровой преобразователь (фильтр)

Скремблер, использующий в кольце обратной связи ЦП, строится по структурной схеме, представленной на рис. 6.29, а. Здесь x_j — последовательность символов ГС; y_j — символы скремблированной последовательности, посылаемой в канал связи; z_j — выходная последовательность ЦП, на вход которого подается y_j . При этом y_j получается посимвольным суммированием по mod 2 последовательностей x_j и z_j :

$$y_j = x_j \oplus z_j. \tag{6.66}$$

В месте приема дескремблирование осуществляется с помощью ЦП, аналогичного использованному на передаче (рис. 6.29, б). Действительно,

$$y_j \oplus z_j = (x_j \oplus z_j) \oplus z_j = x_j \oplus (z_j \oplus z_j) = x_j. \tag{6.67}$$

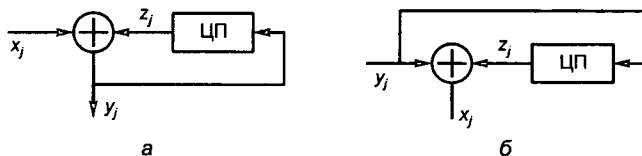


Рис. 6.29. Структурная схема скремблера (а) и дескремблера (б)

Процедура дескремблирования будет соответствовать последнему равенству, если принятый скремблированный сигнал точно равен переданному. Если же при приеме произошла ошибка, то она может породить ошибки дескремблирования все время, в течение которого ошибочный символ находится в элементах памяти ЦП на протяжении ряда передаваемых символов. В результате одна ошибка в канале приводит к нескольким ошибкам дескремблирования, так что в этом варианте ошибки размножаются. Для примера отметим, что скремблирование с использованием ЦП, стандартизованное для сотовой связи GSM, приводит к размножению числа ошибок в три-четыре раза, но не требует специальной синхронизации.

Перейдем к описанию принципа работы СЦС. Пусть рассматривается ГС с длиной цикла $N_{\text{ц}}$ символов, из которых r принадлежат синхрогруппе. Примем для определенности, что последние следуют подряд (сосредоточенная синхрогруппа) и размещены в начале каждого цикла.

СЦС начинает наблюдение принимаемого ГС с некоторого случайно выбранного символа и от него ведет отсчет номеров вновь поступающих символов. Номера можно представить в виде $j = k + lN_{\text{ц}}$, т.е. характеризовать парой чисел (k, l) , где k принимает значения от 1 до $N_{\text{ц}}$, а l — от 1 до ∞ . Первые символы синхрогрупп имеют в первоначальной нумерации номера (g, l) , где g — некоторое конкретное число (фаза циклов), нахождение которого и является задачей СЦС.

СЦС определяет фазу циклов в результате поиска, проверяя для каждого очередного номера k , совпадает он с g или нет. Остановившись в результате поиска на том или ином значении g , СЦС переходит в режим контроля наличия циклового синхронизма, который необходим в связи с тем, что по различным причинам достигнутый синхронизм может со временем нарушиться либо может оказаться, что он был зафиксирован ошибочно.

В СЦС решение о наличии синхронизма принимается не в результате обработки одной синхрогруппы, а еще и в результате проверки периодичности ее повторения на некотором отрезке ГС. Объединение всех синхрогрупп данного отрезка назовем синхросигналом. В СЦС может использоваться параллельный, последовательный или комбинированный (параллельно-последовательный) способ обнаружения синхросигнала. Параллельный анализ считается более эффективным. Однако он требует предварительного запоминания весьма протяженного отрезка ГС, на котором размещается синхросигнал. Чаще используют комбинированный способ, при котором обнаружение отдельных синхрогрупп осуществляется путем параллельного анализа и проверка периодичности их следования — путем последовательного.

Параллельный обнаружитель синхрогрупп представляет собой регистр сдвига с отводами, содержащий r элементов памяти, работающий с тактовой частотой подаваемого на его вход ГС (рис. 6.30). Отводы выполнены от каждого элемента регистра, и в каждом отводе сигнал умножается на соответствующий символ синхрогруппы a_j . После умножения на всех отводах появляется единица в том и только в том случае, когда в регистр записана комбинация, совпадающая с синхрогруппой. Схема совпадения на каждом такте группового сигнала фиксирует совпадение текущего содержимого регистра с синхрогруппой, причем в зависимости от длины синхрогруппы и вероятности ошибки в канале совпадением может считаться как полное совпадение, так и частичное (в $r - \rho$ символах, при $\rho \ll r$).

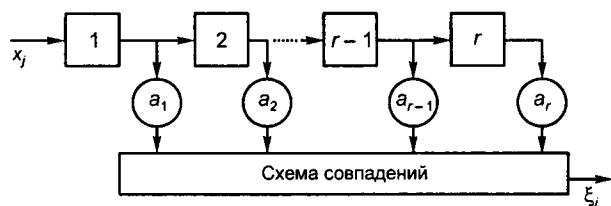


Рис. 6.30. Обнаружитель синхрогрупп

На выходе схемы совпадений образуется последовательность $+1$ (совпадение) и -1 (несовпадение) с тактовой частотой F_r . Будем именовать ее последовательностью совпадений s_j . Последняя подается на устройство проверки периодичности повторения синхрогрупп, проверяющее наличие в ней периодического появления $+1$ через $N_{\text{ц}}$ символов. При поиске синхронизма устройство проверок работает вначале (до появления в s_j первой $+1$) в режиме ожидания. При появлении первой $+1$ (например, при $k = k_1$) схема переходит в режим проверок. Проверяется наличие $+1$ на одноименных позициях последовательности циклов с номерами (k_l, l) для $l = 1, 2, \dots$. Это означает проверку гипотезы $k = k_1$. На некотором цикле при $l = l_1$ проверка заканчивается, и гипотеза либо подтверждается (тогда СЦС переходит в режим контроля), либо отвергается. В последнем случае устройство проверок периодичности вновь переводится в режим ожидания и с появлением в последовательности совпадений очередной $+1$ описанная процедура проверки периодичности повторяется вновь.

Остановимся еще на типовых решающих правилах в устройствах проверки периодичности. Достаточно широкий класс таких правил реализуются двухпороговыми реверсивными счетчиками, на вход которых подается последовательность символов, выделенных с одноименных позиций циклов из последовательности совпадений. Если очередной такой символ $+1$, то счетчик добавляет к накопленному в нем числу некоторую величину d_+ (положительный шаг), если -1 , то вычитает из него d_- (отрицательный шаг). Если накопленное в счетнике число достигает верхнего порога $D_{\text{в}}$, то принимается решение о правильности проверяемой гипотезы, если достигается нижний порог $D_{\text{н}}$, то гипотеза отвергается.

Важно отметить, что описанное решающее правило учитывает достаточно высокую вероятность ложного набора синхрогрупп. Если проверяемая гипотеза не верна, то число $+1$ на выделенных позициях должно быть все-таки существенно меньше числа -1 , так что по истечении достаточного времени, как правило, окажется достигнутым нижний порог. Если же гипотеза в действительности справедлива, то даже при необнаружении отдельных синхрогрупп из-за ошибок демодуляции все равно на выделенных позициях будет существенно больше $+1$ чем -1 , так что чаще будет достигаться верхний порог. Из этого качественного описания становится ясным, что рациональный выбор порогов и шагов позволяет оптимизировать СЦС в зависимости от условий работы конкретной радиолинии и параметров системы временного уплотнения.

Из изложенного ясно, что время достижения синхронизма, являющееся одной из основных характеристик СЦС, достаточно велико и измеряется большим числом циклов временного уплотнения. Как правило, СЦС является наиболее инерционной из всех систем синхронизации, используемых в приемниках, так что именно она определяет полное время вхождения приемника в синхронизм.

В режиме контроля продолжается работа системы проверки периодичности появления синхрогрупп, однако значения шагов и порогов меняются. Изменения направлены на повышение инерционности принятия системой решений, что целесообразно в связи с повышением степени их надежности, поскольку априори потеря циклового синхронизма после его достижения достаточно маловероятна. С другой стороны, в некоторых системах связи, особенно подвижной, из-за замираний и затенений на трассе распространения радиоволн

наблюдаются достаточно частые нарушения связи, которые можно трактовать как обрывы радиолинии. Если не принимать специальные меры, то при таких обрывах цикловая синхронизация, как правило, нарушается. По этой причине, с одной стороны, приходится применять в приемнике специальные меры, обеспечивающие обрывоустойчивость приема, а с другой — улучшать характеристики СЦС и, прежде всего, время ее вхождения в синхронизм.

6.3.3. Алгоритмы асинхронного ввода

Основным устройством, используемым в различных системах асинхронного ввода/вывода, является так называемая *эластичная память* (ЭП). Она представляет собой оперативную память, объем которой может варьироваться в процессе эксплуатации (рис. 6.31). В этом устройстве входной цифровой видеосигнал подается на регистр сдвига с отводами. Временные положения тактовых точек сигнала определяются СТС и подаются в качестве тактов записи (тактовая частота F_z) на элементы памяти регистра. Считывание информации с регистра памяти осуществляется через переключатель П, обеспечивающий возможность считывания с любого элемента памяти регистра. Такт считывания (тактовая частота F_c) задается от отдельного источника, никак не связанного с формирователем тактовой частоты входного сигнала. ЭП содержит также устройство управления, обеспечивающее переключение переключателя П.

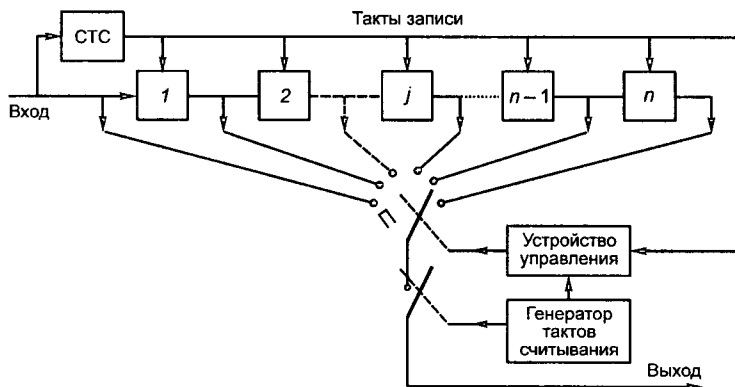


Рис. 6.31. Структура эластичной памяти

Если частоты F_z и F_c не совпадают строго, то при неизменном положении переключателя всегда раньше или позже в выходном сигнале по отношению к входному будут появляться вставки. Однако их можно избежать, переключая П. Действительно, если очередное считывание привело бы к повторению считанного на предыдущем такте символа, следует переместить П на один шаг влево и тем самым обеспечить считывание нового «более свежего» символа. Наоборот, если очередное считывание привело бы к пропуску еще не считанного символа, следует переместить П на шаг вправо, что позволит «догнать» пропущенный символ. Если же данное считывание не приведет ни к одному из этих двух вариантов, то переключатель следует оставить на месте.

Каким образом устройство управления может идентифицировать перечисленные случаи? Это легко сделать, сопоставляя моменты записи и считывания в соответствии с приведенным ниже алгоритмом. При наступлении очередного момента считывания, устройство управления проверяет два следующих обстоятельства:

- является ли данный момент считывания первым на том тактовом интервале записи, на который он пришелся;
- были ли моменты считывания на предыдущем тактовом интервале записи.

Если ответ на второй вопрос оказывается отрицательным, то это означает, что на предыдущем такте был пропущен символ сообщения. Этот символ еще хранится в памяти, но в ячейке с номером $j + 1$, где j — номер ячейки, с которой в данный момент осуществляется считывание. Поэтому, переводя Π в положение $j + 1$ (на шаг вправо), исключим грозившую нам отрицательную вставку.

При положительном ответе вернемся к первому вопросу. Если до данного момента считывания на соответствующем тактовом интервале записи другого момента считывания не было, то это значит, что соответствующий символ сообщения будет считываться впервые, вследствие чего это считывание не приведет к вставке и его можно осуществлять. Если же это считывание будет уже вторым на данном такте, то его осуществлять нельзя, ибо оно привело бы к вставке (в данном случае положительной). Для парирования этой возможности следует переключить Π на шаг влево, после чего осуществить считывание. В результате будет считан очередной еще не транслировавшийся символ сообщения, а вставка будет предотвращена.

Таким образом, использование ЭП при описанном алгоритме управления действительно позволяет избегать вставок. Обратим внимание на то обстоятельство, что такое устранение вставок оказывается возможным лишь до тех пор, пока текущее положение переключателя не достигнет одного из двух крайних положений. Таким образом, описанное устройство само по себе оказывается работоспособным лишь в течение ограниченного интервала времени. Когда речь идет о связи с определенной максимальной длительностью сеанса, такое устройство оказывается достаточным, если только объем ЭП соответствует этой максимальной длительности. В начале сеанса тем или иным способом устанавливают размещаемую на передающей стороне радиолинии ЭП в исходное состояние, после чего она обеспечит отсутствие вставок в течение всего сеанса связи.

Остановимся несколько подробнее на анализе соответствующих процессов. *Во-первых*, всегда желательно, чтобы F_3 и F_c были бы как можно ближе друг к другу. При этом вставки оказываются достаточно редкими, так что при том же объеме памяти ЭП n допустимая длительность сеанса оказывается максимальной. Различают два случая:

- одностороннее расхождение частот (ОРЧ), когда всегда заведомо $F_c > F_3$ или $F_c < F_3$;
- двухстороннее расхождение частот (ДРЧ), когда номинальные значения F_c и F_3 выбираются равными и остается заранее неизвестным, какая из этих частот больше.

Для ДРЧ начальное положение переключателя в ЭП должно быть близким к $n/2$, а для ОРЧ должно составлять 0 (при $F_3 > F_c$) или n (при $F_3 < F_c$).

Если модуль максимального относительного расхождения частот записи и считывания не превосходит δ , скорость передачи информации C , а максимальная длительность сеанса составляет T_c , то должно выполняться условие

$$n > \delta C T_c \text{ для ОРЧ и } n > 2\delta C T_c \text{ для ДРЧ.} \quad (6.68)$$

Не следует думать, что это соотношение указывает на то, что в варианте ДРЧ требуется больший объем памяти. Дело в том, что при одинаковых стабильностях генераторов значения δ для случаев ОРЧ и ДРЧ оказываются различными. Например, пусть относительная стабильность частот записи и считывания одинакова и составляет δ_0 , так что модуль разности этих частот для ДРЧ составляет $\delta = \delta_0$. При той же стабильности генераторов, для того, чтобы гарантировать режим ОРЧ, придется сдвинуть частоту считывания по номиналу относительно частоты записи, по крайней мере, на $2\delta_0$, так что $\delta > 4\delta_0$. Таким образом, требуемый объем памяти для ОРЧ n оказывается по крайней мере в два раза больше, чем при ДРЧ.

Отметим теперь, что относительные нестабильности современных генераторов записи и считывания имеют порядок 10^{-6} – 10^{-7} . Даже с учетом еще встречающегося использования в качестве источников информации устаревшей аппаратуры величина δ не опускается ниже 10^{-5} . По этой причине требуемые объемы ЭП оказываются сравнительно небольшо-

ми. Пусть, например, речь идет о цифровой телефонной связи. Пусть $C = 64$ кбит/с и $T_C = 20$ мин = 1200 с. При $\delta = 10^{-5}$ и ДРЧ из (6.68) найдем для этого случая $n = 1536$, что для современных цифровых устройств не приводит к каким-либо осложнениям. Если же стабильность частоты повышается хотя бы на порядок, то требуемый объем памяти еще в 10 раз сокращается.

Тем не менее описанный вариант асинхронного кодирования оказывается непригодным, если длительность сеанса связи не ограничена. В этом случае чаще всего используется метод асинхронного ввода, именуемый *стаффингом*. Характерной особенностью этого алгоритма является то, что для передачи каждого сообщения применяются два синхронных переносчика: основной и вспомогательный. *Основной переносчик* имеет тактовую частоту F_c , близкую к тактовой частоте сообщения F_s , а вспомогательный — гораздо меньшую тактовую частоту F_n . Основным элементом асинхронного кодера и здесь является устройство ЭП, но с объемом памяти несколько бит. Запись в ЭП производится с тактовой частотой F_s , а считывание с частотой F_c . Поскольку эти частоты не совпадают, сигнал на выходе ЭП содержит вставки (положительные, отрицательные или и те, и другие).

Идея метода стаффинга заключается в том, чтобы, фиксируя на передающей стороне вставки, передавать сведения о них на приемную сторону с использованием вспомогательного переносчика. При этом все основные и вспомогательные последовательности, несущие информацию обо всех уплотняемых сообщениях, являются синхронными, потому могут быть объединены в единый групповой сигнал методом синхронного уплотнения, как это описано в предыдущем параграфе. На приемной стороне сначала производится синхронное разделение, а затем каждая пара — основная и вспомогательная последовательности — подается на свое устройство асинхронного вывода, в котором вначале в соответствии с полученной из вспомогательной последовательности информации о вставках устраняются вставки в основной последовательности. Однако в результате данной процедуры равномерная тактовая структура последовательности нарушается: после устранения положительных вставок появляются такты, не несущие информацию, а после устранения отрицательных вставок за один такт приходится передавать два бита сообщения. Такая «неравномерная» последовательность, как правило, не может быть отдана абоненту, поскольку он не сможет ее правильно воспринять; абоненту сообщение должно быть передано в той форме, в которой оно поступило на передающий конец радиолинии. Поэтому после устранения вставок приходится осуществлять еще одну процедуру — восстановление тактовой частоты транслируемого сообщения и перевод основной последовательности без вставок на эту тактовую частоту.

Опишем теперь в общих чертах как реализуются отмеченные выше процедуры асинхронного кодирования и декодирования. Начнем с кодирования. Остановимся на информации о вставках. Для положительных вставок она сводится лишь к указанию места вставки, а для отрицательных — еще и значения пропущенного бита. Можно ожидать, что, если частоты записи и считывания достаточно близки, то вставки появляются достаточно редко и вспомогательная последовательность, несущая информацию о вставках, может иметь сравнительно низкую тактовую частоту. Однако этому противоречит то обстоятельство, что, как ни редко возникают вставки, их временное положение должно указываться с точностью до такта основного переносчика, что требует, по крайней мере, той же тактовой частоты, что и у этого переносчика. Последнее потребовало бы для реализации асинхронного кодирования, по крайней мере, удвоения пропускной способности, что практически совершенно не приемлемо.

В методе стаффинга эта проблема решается тем, что вставки разрешаются не на любом такте основного переносчика, а лишь на достаточно редких, заранее выбранных, периодически повторяющихся его тактах — тактах разрешенных для вставок (ТРВ). Задержка момента возникновения вставок до ТРВ осуществляется устройством ЭП асинхронного кодера. Запись сообщения в эту память осуществляется с тактовой частотой F_s , а считывание с час-

тотой F_c . При необходимости вставки она устраняется сдвигом на шаг переключателя П. При наступлении ТРВ переключатель возвращается в исходное положение, в связи с чем возникает вставка. Информация о ней и передается на вспомогательном переносчике. Если между двумя последовательными ТРВ (данным и предыдущим) необходимость во вставке не возникала, так что положение переключателя П не изменилось, то в данном ТРВ переключатель не перемещается, вставка не возникает. Если частота появления ТРВ выбрана в соответствии с d_0 , то указанный процесс может продолжаться неограниченно долго, причем положение переключателя будет отклоняться от начального не более, чем на шаг.

С другой стороны, относительно каждого ТРВ по вспомогательному каналу должно быть передано всего несколько бит информации: факт наличия или отсутствия вставки в этом ТРВ (1 бит), знак вставки (1 бит) а для отрицательной вставки еще и содержание пропущенного символа (1 бит). Для передачи данной информации имеется время, равное периоду появления ТРВ. Таким образом, тактовая частота вспомогательной последовательности может быть во много раз меньше, чем у основной. В результате в современных устройствах временного уплотнения с асинхронным вводом методом стаффинга полная избыточность, вызванная как введением маркерных последовательностей, так и реализацией асинхронного ввода не превосходит несколько процентов.

Остановимся на реализации процедуры восстановления тактовой частоты сообщения в асинхронном декодере. Формируемая в таком устройстве освобожденная от вставок неравномерная информационная последовательность (НИП) сопровождается также неравномерной тактовой последовательностью (НТП), представляющей собой последовательность коротких импульсов, отмечающих каждый информационный символ. Только опираясь на такую тактовую последовательность, можно однозначно определить информационное содержание НИП.

Восстановление тактовой частоты транслируемого сообщения или «разглаживание» НИП осуществляется с использованием ЭП, входящей в состав асинхронного декодера. В эту память записывается НИП, причем в качестве тактов записи используется НТП. В результате в регистре памяти ЭП оказывается записанным восстановленный отрезок сообщения без вставок. Первоначально тактовая частота считывания из ЭП устанавливается лишь ориентировочно равной истинному значению тактовой частоты транслируемого сообщения в соответствии с имеющимися на приеме априорными данными о ее величине. Устройство управления ЭП обеспечивает считывание без вставок, отмечая одновременно усредненное направление перемещения переключателя П. Если оно происходит в направлении к началу регистра сдвига, то это означает, что выбранное значение частоты считывания меньше истинного значения тактовой частоты сообщения и ее увеличивают на некоторую заранее выбранную величину, называемую *шагом регулирования*. Если движение П происходит в основном в направлении к началу регистра, то частоту считывания уменьшают на этот шаг. По окончании переходного процесса частота считывания становится приблизительно равной тактовой частоте сообщения, совершая лишь малые колебания вокруг этой величины в пределах, определяемых шагом регулирования.

В заключение настоящего раздела остановимся на задачах асинхронного ввода, возникающих при пакетном методе передачи, используемом при радиоплотнении.

Опишем вначале более подробно, чем это было сделано в 6.3.1, организацию данного режима работы. В этом случае все работающие на общей несущей частоте радиостанции периодически передают пакеты символов, отображающих транслируемые ими сообщения. Период повторения пакетов каждой из станций $T_{ц}$, именуемый также циклом, задается одной из этих станций, которая назначена ведущей в данной сети. Что касается тактовых частот передачи символов внутри пакета, то они определяются каждой станцией для своих пакетов и не совпадают строго для различных станций (обозначим их через F_j , где j — номер станции).

Как отмечалось выше, в системе создается ССП, обеспечивающая синхронизацию излучений отдельных станций, исключая наложение друг на друга излучаемых ими пакетов. Последнее означает, в частности, что для всех станций устанавливается одна и та же длительность цикла $T_{\text{ц}}$. ССП может обеспечить синхронизацию временных положений пакетов различных станций лишь с известной точностью. Для того, чтобы в этих условиях гарантировать отсутствие наложения первых символов данного пакета на последние символы предыдущего, между пакетами предусматривается защитный интервал протяженностью в несколько бит; в пределах данного интервала временное положение пакетов может колебаться в зависимости от различных причин, носящих случайный характер.

Что касается структуры пакета, то она нормируется специальными стандартами, распространяющимися на те или иные системы связи, и включает информационные и служебные блоки. Начинается пакет со служебного блока, предназначенного для синхронизации приемника по несущей и такту. Число бит в этом блоке нормируется так, чтобы оно было достаточным для гарантированного достижения синхронизма. Далее следуют различные служебные комбинации, одна из которых является стартовой. Затем передается информационный блок, завершающийся в общем случае служебными комбинациями, одна из которых рассматривается как стоповая. Заметим, что различные служебные комбинации могут располагаться на известных позициях и внутри информационного блока.

Будем различать: а) случай фиксированных блоков (число бит в каждом из блоков пакета заранее стандартизуется); б) случай нефиксированных блоков (объем информационного блока может в известных пределах варьироваться от пакета к пакету).

В случае фиксированных блоков число информационных бит, передаваемых станцией в единицу времени, составляет $C_j = N_{\text{и}}/T_{\text{ц}}$, где $N_{\text{и}}$ — установленное число бит в информационном блоке пакета. Таким образом, точное значение скорости передачи станцией информационных символов определяется длительностью цикла, задаваемого ведущей станцией. Поэтому данный вариант без дополнительных преобразований может быть использован только в тех случаях, когда источник сообщений допускает управление тактовой частотой сообщения со стороны станции. Такая ситуация наблюдается, например, когда передаче подлежат аналоговые сообщения, причем их преобразование в цифровую форму осуществляется на самих станциях. Тогда тактовая частота сообщения определяется оборудованием станции. Аналогичная ситуация имеет место при получении цифровой информации от компьютера по синхронному стыку, когда тактовая частота выдаваемого сообщения задается по тому же стыку.

Если же тактовая частота сообщения станцией не управляется, то целесообразно использовать вариант нефиксированных блоков, когда асинхронный ввод сообщения обеспечивается за счет вариации числа бит $N_{\text{и}}$ в информационных блоках пакетов на небольшую величину (в пределах нескольких бит). Тогда значения $N_{\text{и}}$ отдельных пакетов данной станции могут быть выбраны так, чтобы при любых возможных значениях $T_{\text{ц}}$ число бит сообщения, переданных в единицу времени, точно равнялось бы числу бит, поступающих от источника. Для того, чтобы при нефиксированных блоках правильно определять границы информационных потоков, используются стартовые и стоповые комбинации, упоминавшиеся при описании структуры пакетов. По этой причине описанный способ асинхронного ввода получил название *стартстопного*. Отметим, что стартовая комбинация необходима и при фиксированной структуре блоков, поскольку начало содержательных блоков все равно необходимо обозначить. А вот в стоповой комбинации нет необходимости: ее может заменить простой подсчет символов пакета.

В заключение отметим, что пакетное радиоуплотнение с фиксированными блоками используется в сетях сотовой связи стандарта GSM.

Глава 7

ШИРОКОПОЛОСНЫЕ СИГНАЛЫ В СИСТЕМАХ СВЯЗИ

7.1. Методы широкополосной передачи

Свое название широкополосные системы связи получили вследствие того, что полоса, занимаемая используемыми в них сигналами, намного шире полосы, необходимой для передачи непосредственно информации. Одной из первых таких систем, по-видимому, являлась разработанная в конце 1950-х гг. система «Рейк» [7.1]. В этой системе за счет использования метода широкополосной передачи удалось обеспечить устойчивую связь в условиях многолучевого распространения. Методы широкополосной передачи позволили осуществить разделение нескольких лучей с различным запаздыванием и тем самым устранить эффект замирания сигналов, вызванный многолучевым распространением. В специальных системах методы широкополосной передачи позволяют организовать устойчивую передачу информации в условиях действия преднамеренных помех, мощность которых на входе приемника может превышать мощность полезных сигналов в сотни и тысячи раз. Кроме того, в таких системах использование методов широкополосной передачи позволяет затруднить средствам радиоразведки обнаружение факта передачи, т.е. повысить ее скрытность. В сотовых и спутниковых системах связи (см. гл. 6) методы широкополосной передачи позволяют обеспечить одновременную работу многих пользователей в общей полосе частот, т.е. реализовать метод многостанционного доступа, основанный на разделении сигналов по форме (Code Division Multiple Access, CDMA).

В системах радиолокации использование методов широкополосной передачи позволяет повысить точность измерения дальности до цели при прочих равных условиях, а также преодолеть известное противоречие между дальностью действия локатора и его разрешающей способностью.

Среди методов широкополосной передачи в цифровых системах связи наибольшее распространение получили два метода. Первый метод расширения спектра основан на использовании псевдослучайных последовательностей (ПСП). Такие сигналы обычно называют широкополосными (ШПС), или шумоподобными. Наиболее полное изложение теории и техники шумоподобных сигналов можно найти в работах Л. Е. Варакина [7.2–7.4].

Углубленная функциональная схема (модель) цифровой системы связи с ШПС приведена на рис. 7.1.

Функции, выполняемые кодером и декодером в этой модели, обсуждались в предыдущих разделах, хотя использование кодирования, как будет ясно из дальнейшего, в широкополосных системах имеет определенные особенности. Генераторы ПСП на передающей и приемной сторонах идентичны. Именно они сначала применяются для расширения спектра

передаваемых по каналу связи сигналов, а затем перед демодуляцией для его сжатия. Для расширения спектра в такой схеме применяют фазовую манипуляцию, а получаемые при этом сигналы, как это отмечено выше, нередко называют ФМ ШПС. Информационная манипуляция также фазовая, хотя возможна и произвольная. В модуляторе сначала осуществляется перемножение кодированных символов с ПСП (расширение спектра), а затем непосредственно фазовая манипуляция.

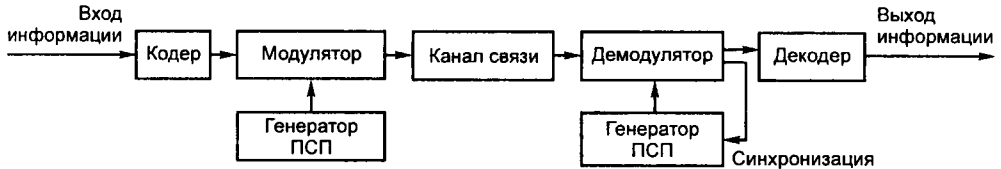


Рис. 7.1. Модель цифровой системы связи с ШПС

Второй часто используемый метод широкополосной передачи основан на псевдослучайной перестройке рабочей частоты сигнала (ППРЧ).

Укрупненная функциональная схема (модель) цифровой системы связи с ППРЧ приведена на рис. 7.2.

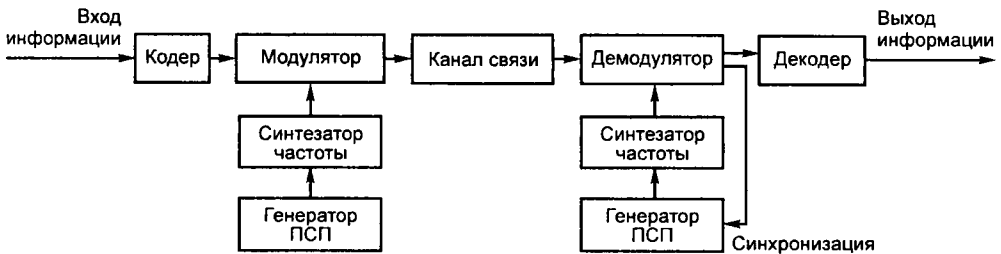


Рис. 7.2. Модель цифровой системы связи с ППРЧ

Отличаются две схемы тем, что во второй расширение спектра осуществляется не за счет перемножения кодированной информации с ПСП, а за счет вырабатываемой синтезатором и перестраиваемой по псевдослучайному закону рабочей (несущей) частоты модулятора.

На приемной стороне производится обратное преобразование, что приводит к сжатию спектра перед демодуляцией. При ППРЧ информационная манипуляция также может быть произвольной, хотя следует отметить, что в этом случае в моменты смены частот могут наблюдаться случайные скачки начальной фазы несущей, поэтому может потребоваться некогерентная демодуляция, а это заметно снижает эффективность кодирования.

Отметим, что сжатие спектра принимаемого сигнала происходит только в том случае, когда система синхронизации приемника совместит по времени и по частоте принимаемый сигнал и опорный сигнал, вырабатываемый местным генератором ПСП.

Рассмотрим каждый из указанных методов широкополосной цифровой передачи более подробно.

7.2. Шумоподобные сигналы и псевдослучайные последовательности

Как следует из модели цифровой системы связи с ШПС, расширение спектра передаваемых в канал связи сигналов осуществляется за счет перемножения или, что то же, сложения по mod 2 кодированной (или некодированной, если кодирование не используется) информации с ПСП.

Псевдослучайная последовательность $a(k)$ и комплексная огибающая $U(t)$ образованно на ее основе ШПС связаны следующим соотношением:

$$U(t) = \sum_{k=0}^{N-1} u_0(t - k\tau_0) \exp\{i\pi a(k)\}, \quad (7.1)$$

где $u_0(t)$ и τ_0 — огибающая и длительность элементарного символа сигнала; N — длина, т.е. количество элементарных символов ПСП, укладываемых на длительности кодированного символа. Для упрощения изложения будем полагать, что элементарные символы имеют прямоугольную форму и амплитуду, равную единице.

Наибольшее распространение в действующих системах получили двоичные ПСП, символы которых принимают значения 0 и 1. Как уже отмечалось в гл. 6, ШПС, используемые в широкополосных системах, характеризуются параметром, который называют *базой сигнала* B . База определяет степень расширения спектра сигнала и количественно определяется числом символов ПСП, укладываемых на длительности информационного (или кодированного) символа или, что то же, произведением полосы F , занимаемой спектром ШПС, на длительность информационного (или кодированного) символа T . Для ШПС $B \gg 1$, в то время как в системах без расширения спектра $B \approx 1$, а сигналы называют *простыми*, или *узкополосными*.

Основное требование, которому должны удовлетворять ПСП в широкополосных системах, вытекает из их названия. Это псевдослучайность, или шумоподобность. Такие свойства ПСП, например, как «хорошая» автокорреляционная функция (АКФ), т.е. с малыми боковыми лепестками, или наиболее равномерный амплитудный спектр, являются производными от их псевдослучайности. Такими же свойствами обладает достаточно длинная реализация БГШ. Ее АКФ представляется в виде δ -функции Дирака, энергетический спектр — равномерный.

Апериодическая и периодическая АКФ комплексной огибающей ШПС, построенного с использованием двоичной ПСП $a(k)$, в дискретных точках $\tau = k\tau_0$ могут быть представлены в следующем виде:

$$R(m) = \begin{cases} \frac{1}{N} \sum_{k=0}^{N-m-1} e^{i\pi[a(k)-a(k+m)]} = \frac{1}{N} \sum_{k=0}^{N-m-1} d(k)d(k+m), & \text{апериодическая;} \\ \frac{1}{N} \sum_{k=0}^{N-1} e^{i\pi[a(k)-a(k+m)]} = \frac{1}{N} \sum_{k=0}^{N-1} d(k)d(k+m), & \text{периодическая,} \end{cases} \quad (7.2)$$

$$\text{где последовательность } d(k) = \begin{cases} +1, & \text{если } a(k) = 0; \\ -1, & \text{если } a(k) = 1, \end{cases} \quad (7.3)$$

а сумма $(k+m)$ вычисляется по mod N .

Рассмотрим, каким требованиям должны удовлетворять ПСП, чтобы их можно было бы использовать в качестве основы для построения ШПС.

7.2.1. Последовательности Баркера

Первоначально критерием для поиска «хороших» последовательностей являлся минимум боковых лепестков их апериодических АКФ. В наилучшей степени этому критерию удовлетворяли двоичные последовательности (коды) Баркера. Было найдено только шесть последовательностей Баркера длины $N = 3, 4, 5, 7, 11,$ и 13 . Сами последовательности Баркера и значения боковых лепестков их АКФ приведены в табл. 7.1.

Таблица 7.1. Последовательности Баркера

N	$d(k)$													$R(2l)$	
	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$	$k=11$	$k=12$		
3	1	1	-1												-1/3
4	1	1	1	-1											0
5	1	1	1	-1	1										1/5
7	1	1	1	-1	-1	1	-1								-1/7
11	1	1	1	-1	-1	-1	1	-1	-1	1	-1				-1/11
13	1	1	1	1	1	-1	-1	1	1	-1	1	-1	1		1/13

Эти последовательности состоят из символов $d(k) = 1$ и -1 и в соответствии с (7.2) для нечетных N имеют апериодическую АКФ вида

$$R(m) = \frac{1}{N} \sum_{k=0}^{N-m-1} d(k)d(k+m) = \begin{cases} 1, & \text{для } m=0; \\ 0, & \text{для } m=2l+1, \text{ где } l=0, 1, \dots, \frac{N-1}{2}; \\ \pm 1/N, & \text{для } m=2l. \end{cases} \quad (7.4)$$

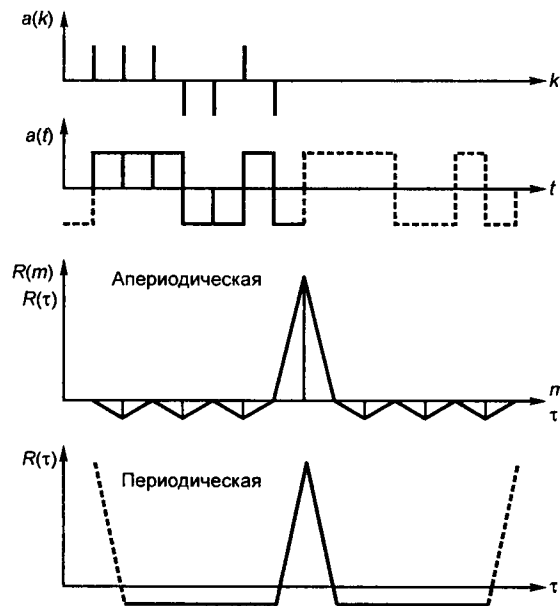


Рис. 7.3. Последовательность Баркера длины $N = 7$ и ее АКФ

Из выражения (7.4) следует, что уровень максимальных боковых лепестков АКФ последовательностей Баркера равен $(+1/N)$ или $(-1/N)$. Последовательность Баркера для $N = 7$, фазоманипулированный видеосигнал, построенный на ее основе, апериодическая и периодическая АКФ этого видеосигнала представлены на рис. 7.3.

Поиски последовательностей Баркера большей длины успехом не увенчались. Из-за относительно малой длины последовательности Баркера широкого применения в действующих широкополосных системах практически не нашли.

7.2.2. M -последовательности

Наиболее широкое применение в широкополосных системах связи нашли так называемые M -последовательности. Как правило, используются двоичные M -по-

следовательности, символы которых $a(k)$ и $d(k)$ с учетом (7.3) принимают значения $a(k)$ 1 и 0 и $d(k)$ соответственно -1 и 1 . Такие последовательности обладают следующими свойствами [7.2]:

- 1) M -последовательности являются периодическими с периодом $N = 2^n - 1$ символов, где n — произвольное целое положительное число;
- 2) количество символов, принимающих значение единица, на длине одного периода M -последовательности равно 2^{n-1} , что на единицу больше, чем количество символов, принимающих значение нуль;
- 3) любые комбинации символов длины n на длине одного периода M -последовательности за исключением комбинации из n нулей встречаются не более одного раза. Комбинация из n нулей является запрещенной, на ее основе может генерироваться только последовательность из одних нулей;
- 4) сумма по mod 2 любой M -последовательности с ее произвольным циклическим сдвигом также является M -последовательностью;
- 5) периодическая АКФ любой M -последовательности с учетом (7.3) имеет постоянный уровень боковых лепестков, равный $(-1/N)$.

Уровень максимальных боковых лепестков аperiodической АКФ примерно составляет $1/\sqrt{N}$.

Формируются M -последовательности многотактными линейными фильтрами в виде регистров сдвига с обратной связью (см. 4.4). Для формирования M -последовательности с периодом $N = 2^n - 1$ может использоваться регистр сдвига длины n . Структура обратных связей регистра определяются неприводимым, примитивным многочленом степени n . Определение и таблицу неприводимых, примитивных многочленов для $n \leq 40$ можно найти в [7.5]. Схема фильтра, построенного на основе многочлена $G(x) = X^5 + X^2 + 1$, генерирующего M -последовательность длины $N = 31$, представлена на рис. 7.4. Начальные n символов M -последовательности определяются исходным состоянием ячеек регистра. M -последовательность может генерироваться из любого исходного состояния регистра, отличного от всех нулей. Как уже отмечалось выше, комбинация из n нулей является запрещенной. С поступлением на тактовый вход регистра очередного тактового импульса состояния ячеек регистра последовательно меняется (на данной схеме слева направо).

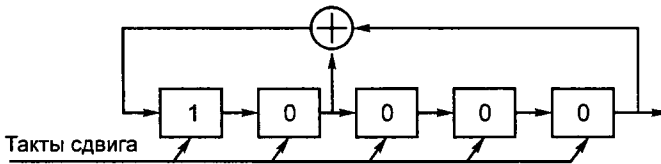


Рис. 7.4. Генератор M -последовательности длины $N = 31$

M -последовательность (один период), генерируемая таким регистром из начального состояния регистра 10 000, имеет следующий вид:

0 0 0 0 1 0 0 1 0 1 1 0 0 1 1 1 1 1 0 0 0 1 1 0 1 1 1 0 1 0 1

Отметим, что разным разрешенным состояниям регистра соответствуют различные циклические сдвиги одной M -последовательности.

Приведенная схема формирователя позволяет проиллюстрировать выполнение некоторых из перечисленных выше свойств M -последовательности. Так, зная, что период M -последовательности $N = 2^n - 1$, можно легко представить выполнение свойства 3. Действительно,

если предположить, что произвольная комбинация из n символов появится менее, чем через период $N = 2^n - 1$, это будет означать, что регистр из одних и тех же начальных условий генерирует разные последовательности, что противоречит физическому смыслу.

Количество различных M -последовательностей с периодом N равно отношению $\varphi(N)/n$, где $\varphi(N)$ — функция Эйлера, равная количеству чисел в ряду $1, 2, \dots, N-1$ взаимно простых с числом N .

Последовательности, периодические АКФ которых имеют уровень максимальных боковых лепестков, равный $(-1/N)$, называются *минимаксными*. Таким образом, M -последовательности являются минимаксными последовательностями.

Наряду с M -последовательностями к минимаксным относятся ПСП символов Лежандра, последовательности Якоби и Холла [7.2]. Однако эти последовательности широкого применения в действующих системах не нашли, поэтому ограничимся только их перечислением.

7.2.3. Ансамбли псевдослучайных последовательностей

Ансамблем последовательностей (видеосигналов) называется некоторое множество последовательностей, определяемое общим правилом построения [7.3].

В спутниковых и сотовых широкополосных системах связи каждому абоненту (постоянно или только на время сеанса связи) присваивается некоторая ПСП, которая может служить некоторым аналогом его адреса. Поэтому для систем связи с большим числом абонентов должен быть выбран и ансамбль различных ПСП большого объема, т.е. число ПСП, входящих в его состав, должно быть достаточно большим ($M \gg 1$). Сигнал, соответствующий присвоенной данному абоненту ПСП, является для него полезным, сигналы других абонентов оказываются для него мешающими, т.е. взаимной (системной) помехой. Поэтому ансамбли ПСП для систем связи следует выбирать по такому критерию, который обеспечит минимум взаимных помех при прочих равных условиях.

Как показано в работе [7.3], в случае, когда уровни всех мешающих сигналов примерно одинаковы, и согласно центральной предельной теореме имеет место нормализация распределения суммарной системной помехи, образованной мешающими сигналами, в основу выбора системы ПСП для широкополосных систем связи может быть положен критерий минимума дисперсии боковых лепестков их взаимокорреляционных функций (ВКФ). В случае равенства дисперсий ансамбль ПСП целесообразно выбирать по минимуму разброса дисперсий отдельных ВКФ, это правило обеспечивает максимум помехоустойчивости приема информации при прочих равных условиях. Однако нередко в системах связи без механизма выравнивания мощности уровни мешающих сигналов от отдельных абонентов могут сильно отличаться. В основном это объясняется различным местоположением абонентов. Распределение взаимной помехи может заметно отличаться от нормального (гауссовского) и ансамбль ПСП следует выбирать по минимуму максимальных боковых лепестков ВКФ. Вычисление всех ВКФ ансамбля ПСП достаточно большого объема ($M > 100$) представляет собой очень трудоемкую задачу, решение которой к тому же может оказаться бесполезным, если выбор для анализа ансамбля ПСП произведен неудачно. Поэтому представляет интерес рассмотреть методы построения ансамблей «хороших» ПСП, которые обеспечивают заведомо большую вероятность получения желаемого результата.

Ансамбли последовательностей Уолша. Правило построения ансамблей последовательностей (функций) Уолша приведено в 4.3, где показано их использование в качестве ортогональных m -ичных кодов.

Эти последовательности обладают многими положительными качествами. Они ортогональны, легко формируются современной цифровой техникой. Однако в качестве ПСП в широкополосных системах «прямого» использования они не нашли. Анализ показывает, что дополнительная манипуляция произвольной последовательностью Уолша практически не приводит к расширению спектра сигнала, т.е. база сигнала при этом остается близкой к единице. В результате такой манипуляции в зависимости от номера выбранной последовательности по аналогии с m -ичной частотной манипуляцией на ту или иную величину происходит сдвиг спектра сигнала по частоте [7.3], расширения спектра при этом практически не происходит. Правило построения последовательностей Уолша не вносит в них элемента псевдослучайности и, как следствие этого, корреляционные свойства последовательностей Уолша нельзя считать удовлетворяющими приведенным выше критериям. Уровень максимальных боковых лепестков их АКФ и ВКФ чрезвычайно высок (достигает значений $N - 1$ при длине периода ПСП, равной N), следовательно, и высок разброс дисперсий их боковых лепестков. Но на основе систем последовательностей Уолша можно строить производные ансамбли [7.3] ПСП, которые обладают вполне удовлетворительными корреляционными характеристиками и находят применение в ряде реальных широкополосных систем связи.

Производные ансамбли ПСП, построенные на основе последовательностей Уолша. *Производным* называется ансамбль двоичных последовательностей, который получается в результате посимвольного перемножения (при использовании в качестве элементарных символов $+1$ или -1) или суммирования по $\text{mod } 2$ (при элементарных символах 0 и 1) последовательностей некоторого исходного ансамбля с некоторой последовательностью, называемой *производящей*. Последняя не должна совпадать ни с одной из последовательностей исходного ансамбля. Длина производящей последовательности выбирается равной длине последовательностей в исходном ансамбле. Как показал анализ, производящую ПСП следует выбирать так, чтобы ее аperiodическая АКФ имела низкий уровень максимальных боковых лепестков.

Одним из первых известных производных ансамблей последовательностей был ансамбль, описанный в [7.6]. Здесь в качестве исходного был выбран ансамбль последовательностей Уолша объемом $m = 16$. Роль производящей играла ПСП длины $N = 16$ символов, АКФ которой имела нормированный уровень максимальных боковых лепестков, равный $1/\sqrt{N}$. Исследования корреляционных свойств ансамблей производных последовательностей, построенных на основе последовательностей Уолша длины $N = 16, 32$ и 64 и производящих ПСП соответствующей длины (рис. 7.5), показали, что уровень максимальных нормированных боковых лепестков построенных производных ансамблей последовательностей составляет соответственно $9/16, 12/32$ и $17/64$, что достаточно близко к эмпирической границе $(2-3)/\sqrt{N}$. Это значительно ниже уровней максимальных нормированных боковых лепестков исходных последовательностей Уолша, которые составляют соответственно $15/16, 31/32$ и $63/64$.

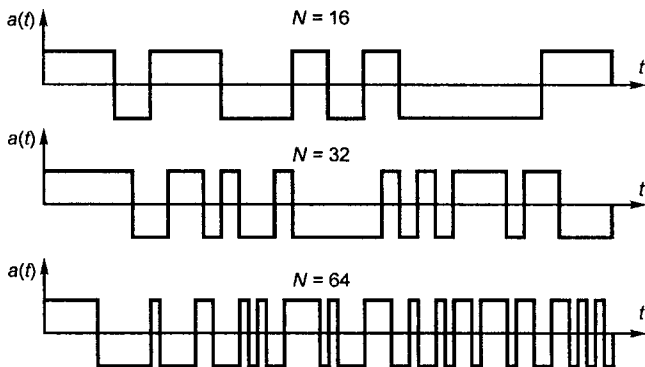


Рис. 7.5. Производящие ПСП длины $N = 16, 32, 64$

Производные ансамбли последовательностей Голда. Для формирования ансамбля двоичных последовательностей Голда в качестве исходной выбирается ансамбль последовательностей в виде циклических сдвигов некоторой двоичной M -последовательности $a(k)$ длины N . В качестве производящей выбирается другая двоичная M -последовательность $b(k)$ той же длины N . Правило выбора этих M -последовательностей $a(k)$ и $b(k)$ предложено Голдом [7.7].

Последовательность производного ансамбля с номером p строится следующим образом:

$$c_p(k) \equiv a(k+p) \oplus b(k), \quad k = \overline{0, N-1}, \quad p = \overline{0, N-1}, \quad (7.5)$$

где знак \oplus означает суммирование по mod 2, а $(k+p)$ вычисляется по mod N . Объем такого ансамбля равен N .

Производный ансамбль последовательностей, построенных на основе двух M -последовательностей длины $N=7$

$$a(k) = 1110010$$

$$b(k) = 1110100,$$

имеет следующий вид:

$$c_0(k) = 0000110$$

$$c_1(k) = 0011011$$

$$c_2(k) = 0100001$$

$$c_3(k) = 1010101$$

$$c_4(k) = 0111100$$

$$c_5(k) = 1101111$$

$$c_6(k) = 1001000. \quad (7.6)$$

Можно показать, что сигналы, соответствующие последовательностям ансамбля (7.5), являются симплексными, т.е. их коэффициент корреляции $r = \frac{-1}{N}$ (см. (4.14)). Выбранная по правилу Голда пара M -последовательностей имеет трехуровневую периодическую ВКФ, т.е.

$$R_{ab}(m) = \frac{1}{N} \sum_{k=0}^{N-1} \exp \{ i\pi [a(k) - b(k+m)] \} \quad (7.7)$$

принимает для каждого m одно из трех возможных значений

$$R_1 = -1/N \text{ или } R_2 = f(n) - 1/N, \text{ или } R_3 = -[f(n) + 1/N],$$

$$\text{где } f(n) = \frac{1}{N} \begin{cases} 2^{(n+1)/2}, & \text{если } n - \text{нечетное;} \\ 2^{(n+2)/2}, & \text{если } n - \text{четное.} \end{cases} \quad (7.8)$$

Поскольку последовательности $a(k)$ и $b(k)$ являются M -последовательностями, то каждая из них обладает циклическим свойством, т.е. сумма по mod 2 любой M -последовательности с ее произвольным циклическим сдвигом также является некоторым циклическим сдвигом этой же M -последовательности.

Используя это свойство, получаем

$$c_p(k) \oplus c_q(k+m) \equiv [a(k) \oplus b(k+p)] \oplus [a(k+m) \oplus b(k+m+q)] \equiv a(k,m) \oplus b(k,m,p,q), \quad (7.9)$$

где $a(k,m) = a(k) \oplus a(k+m)$ и $b(k,m,p,q) = b(k+p) \oplus b(k+m+q)$ — некоторые циклические сдвиги исходных M -последовательностей $a(k)$ и $b(k)$.

Выражение (7.9) фактически служит доказательством того, что и периодические ВКФ последовательностей производного ансамбля (7.6), построенных по правилу Голда, также

являются трехуровневыми со значениями $R_{pq}(m)$, определяемыми выражениями (7.8). При этом боковые лепестки периодических АКФ последовательностей также принимают значения из (7.8).

Последовательности Касами формируются также на основе некоторой исходной M -последовательности. Алгоритм формирования этих ПСП следующий. Выбирается произвольная двоичная M -последовательность $a(k)$ длины $N = 2^n - 1$, где n — четное. Из последовательности $a(k)$ методом децимации (прореживания) через $2^{n/2} + 1$ символов, начиная с произвольной фазы, формируется новая периодическая последовательность $b(k)$ с длиной периода $L = 2^{n/2} - 1$ символов. Например, при $n = 10$ длина исходной M -последовательности $a(k)$ составит $N = 2^n - 1 = 1023$ символа, а длина периода последовательности $b(k)$ равна $L = 31$ символу. Пусть $b_0(k)$ представляет собой периодически повторяющуюся $2^{n/2} + 1$ раза последовательность символов $b(k)$ с номерами 1, 34, 67, ..., 958, 991 из $a(k)$. Последовательность $b_i(k)$, формируемая по аналогии с $b_0(k)$, и представляет собой i -й циклический сдвиг последовательности $b(k)$ также повторяющийся 33 раза. (Здесь $i = 0, 1, \dots, 30$.)

Ансамбль последовательностей Касами $c_i(k)$ объемом $2^{n/2}$ ПСП формируется суммированием по mod 2 символов последовательности $a(k)$ длины $N = 2^n - 1$ и всех, периодически повторяющихся $2^{n/2} + 1$ раз каждая, последовательностей $b_i(k)$. К числу ПСП Касами относится и исходная M -последовательность $a(k)$. В итоге имеем ансамбль последовательностей Касами $c_i(k)$ длины $N = 2^n - 1$ символа и объемом $(2^{n/2})$ ПСП. (Для $n = 10$ это 32 ПСП длины 1023 символа.)

Периодические АКФ и ВКФ последовательностей Касами

$$R_y(m) = \frac{1}{N} \sum_{k=0}^{N-1} \exp\{i\pi \cdot [c_i(k) - c_j(k+m)]\} \quad (7.10)$$

для каждого m принимают одно из трех возможных значений

$$R_1 = -\frac{1}{N} \quad \text{или} \quad R_2 = \frac{(\sqrt{N}-1)}{N}, \quad \text{или} \quad R_3 = \frac{-(\sqrt{N}+1)}{N}.$$

По сравнению со значениями периодических АКФ и ВКФ последовательностей Голда той же длины это примерно в два раза ниже. Однако объем системы последовательностей Касами значительно меньше объема последовательностей Голда (почти в корень квадратный из N).

Сегментные ПСП. *Сегментными* называют ансамбли ПСП, образованные из сегментов (отрезков) M -последовательностей большой длины. Применение таких ПСП в широкополосной системе УКВ радиостанций было отмечено в работе [7.8]. Здесь M -последовательность длины $N = 131\,071$ символов разбивалась на неперекрывающиеся сегменты с длиной $N_0 = 63$ символа каждый. Было получено $L = 2080$ сегментов. Из них методом простого перебора были отобраны примерно 1000 сегментов, нормированные боковые пики АКФ и ВКФ которых не превышали 0,25. При больших значениях N , L , и N_0 выбор сегментов методом простого перебора даже при использовании современных компьютеров может оказаться чрезвычайно трудоемким, а во многих случаях и бесперспективным. Представляет интерес определить какой-либо метод, позволяющий найти аналитические оценки уровня максимальных боковых пиков, и тем самым упростить расчеты на ЦВМ.

Для нахождения такой оценки можно воспользоваться циклическим свойством M -последовательностей: сумма по mod 2 любой M -последовательности $S(k)$ с ее произвольным циклическим сдвигом $S(k+j)$ также является M -последовательностью. Из этого свойства следует, что и сумма по mod 2 двух сегментов M -последовательности является сегментом этой

же M -последовательности, но с другим сдвигом. ВКФ сегментов $s_p(t)$ и $s_q(t)$ в дискретных точках $\tau = m\tau_0$ с учетом преобразования (7.3) имеет следующий вид:

$$R_{pq}(m) = \frac{1}{N_0} \sum_{k=0}^{N_0-m-1} s_p(k)s_q(k+m). \quad (7.11)$$

Используя сформулированное выше циклическое свойство сегментов M -последовательности, из (7.11) получим

$$R_{pq}(m) = \frac{1}{N_0} \sum_{k=0}^{N_0-m-1} s_j(k), \quad (7.12)$$

где $s_j(k)$ — l -й сегмент исходной M -последовательности длины $N_0 - m$.

Таким образом, оценка уровня максимальных боковых пиков АКФ и ВКФ сегментов эквивалентна нахождению значения максимального бокового пика среди всех непериодических АКФ, соответствующих исходной M -последовательности и всех ее циклических сдвигов. Обозначим через $R_j(m)$ АКФ M -последовательности $S(k)$ длины N со сдвигом j :

$$NR_j(m) = \sum_{k=0}^{N-m-1} S(j+k)S(j+k+m) = \sum_{k=0}^{N-m-1} S(l+k). \quad (7.13)$$

При изменении j от 0 до $N-1$ индекс l в (7.13) пробегает все те же значения, что и j , но в ином порядке. Поэтому при таком изменении j сумма в правой части (7.12) дает значения всех боковых пиков непериодических АКФ сегментов длины $N-m$. В свою очередь, изменение m от 0 до $N-1$ определяет все сегменты с длиной, изменяющейся от 1 до N . Это свойство АКФ M -последовательностей и было использовано для вычисления максимума суммы

$$\sum_{k=0}^{N_0-m-1} s_j(k). \quad (7.14)$$

Для $N > 15$ большой объем расчетов показал, что

$$\sum_{k=0}^{N_0-m-1} s_j(k) < \alpha\sqrt{N}, \quad \text{где } \alpha = 1,26. \quad (7.15)$$

Таким образом, используя выражение (7.15), из (7.11) и (7.12) получим верхнюю оценку уровня максимальных боковых пиков непериодических АКФ и ВКФ сегментов M -последовательностей

$$\max R_{pq}(m) < 1,26 \frac{\sqrt{N}}{N_0}. \quad (7.16)$$

Однако оценка (7.16) при вычислениях практически не встречалась. Для большинства рассмотренных M -последовательностей имела место оценка

$$\max R_{pq}(m) \approx \frac{\sqrt{N}}{N_0}, \quad (7.17)$$

т.е. в этом случае $\alpha = 1$.

Приведем примеры расчетов АКФ и ВКФ двух ансамблей последовательностей, являющихся сегментами M -последовательностей длины $N = 255$ (первообразный многочлен $X^8 + X^5 + X^3 + X + 1$) и $N = 511$ (первообразный многочлен $X^9 + X^4 + 1$). Предварительно были определены все значения сумм $\sum_{k=0}^{N_0-m-1} s_j(k)$ для всех возможных сегментов. В результате

были уточнены коэффициенты α . Оказалось, что для выбранной M -последовательности с

$N = 255$ коэффициент $\alpha = 0,81$, а для $N = 511$ $\alpha = 1,06$. Задав значения $R_{pq}(m) = 0,4$ для $N = 255$ и $\alpha = 0,81$ и $R_{pq}(m) = 0,3$ для $N = 5$ и $\alpha = 1,06$, можно найти длины сегментов N_0 , удовлетворяющих этим значениям и количество сегментов L . В результате для $N = 255$ получим $N_0 = 33$; $L = 12$, для $N = 511$ — $N_0 = 80$; $L = 9$. Последующие расчеты непериодических АКФ и ВКФ показали, что реальные значения боковых пиков не превосходят, но весьма близки к оценочным.

7.3. Некоторые особенности использования ШПС в системах связи

7.3.1. Эффективность использования отведенной полосы частот

Еще до недавнего времени использование ШПС в многоканальных спутниковых или сотовых системах подвижной связи считалось нецелесообразным, поскольку при асинхронной работе им свойственны взаимные (системные) помехи, а это приводит к снижению пропускной способности системы, т.е. числа одновременно действующих каналов связи при неизменной скорости передачи по сравнению с системами, использующими простые сигналы и традиционные методы многостанционного доступа, основанные на временном и частотном разделении сигналов различных абонентов при их асинхронной работе. Отметим, что в канале с БГШ наличие взаимных помех при использовании ШПС обязательно предполагает наличие какого-либо энергетического запаса в радиолиниях, т.е. реальное значение E_6/N_0 должно быть больше минимально необходимого для обеспечения заданной вероятности ошибки. Величина такого запаса при низких его значениях будет ограничивать пропускную способность системы с ШПС. При этом пропускная способность системы с простыми сигналами не зависит от энергетического запаса и при прочих равных условиях всегда будет оставаться выше, чем при использовании ШПС.

Относительно взаимных помех следует отметить, что при синхронной работе (когда ШПС формируются от единого опорного генератора, и все они попарно ортогональны) возникновение взаимных помех удастся избежать. Это, например, имеет место в прямых каналах (от базовой станции к мобильным абонентам), описываемой в гл. 11 спутниковой системе «Globalstar», или в сотовых системах связи стандарта CDMA. Однако в обратных каналах этих систем избежать возникновения системных помех практически не удастся.

Однако наличие пауз при телефонных разговорах, а также свойственные системам подвижной связи такие особенности работы, как вызванные многолучевым распространением замирания сигнала доплеровские сдвиги частоты и переменные задержки представляют определенный интерес к использованию ШПС в таких системах. Исследования [7.9, 7.10] показали, что в системах ШПС не только не дают проигрыша, но и обеспечивают значительный выигрыш как по пропускной способности, так и по помехоустойчивости и надежности связи.

Оценим пропускную способность асинхронной системы связи с ШПС. Пусть в выделенной полосе частот F одновременно работает M станций, излучающих ШПС разной формы, но примерно одинаковой мощности S_r . Выравнивание мощностей сигналов в современных системах подвижной связи на входе ретранслятора или базовой станции обеспечивают системы регулировки мощности.

Поскольку все ШПС передаются в общей полосе частот, на входе приемника одного из сигналов остальные $(M-1)$ сигналов оказываются взаимной помехой. Мощность взаимной помехи

$$P_{\text{пом}} = \sum_{i=1}^{M-1} S_i = (M-1)S \text{ при } S_i \approx S. \quad (7.18)$$

Распределение взаимной помехи даже при небольшом числе мешающих ШПС близко к гауссовскому. Спектральная плотность мощности взаимной помехи

$$N_{\text{пом}} = \frac{(M-1)S}{F}. \quad (7.19)$$

Суммарная спектральная плотность мощности взаимных помех и БГШ

$$N_{\Sigma} = \frac{(M-1)S}{F} + N_0. \quad (7.20)$$

Если все станции передают информацию с одинаковой скоростью R , то энергия, приходящаяся на бит передаваемой информации, $E_6 = S/R$. Таким образом, отношение сигнал/суммарная помеха в полосе ШПС F :

$$\frac{E_6}{N_{\Sigma}} = \frac{S/R}{\frac{(M-1)S}{F} + N_0}. \quad (7.21)$$

Здесь E_6/N_{Σ} — минимально допустимое значение отношения сигнал/помеха, при котором обеспечивается требуемое качество приема информации. Из (7.21) легко найти допустимое число одновременно действующих каналов, т.е. пропускную способность системы с ШПС

$$M = \left\lfloor \frac{F}{R} \left(\frac{N_{\Sigma}}{E_6} - \frac{N_0}{E_6} \right) \right\rfloor + 1, \quad (7.22)$$

где $\lfloor A \rfloor$ — целая часть числа A ; $R = 1/T$; T — длительность информационного бита, $B = F/R = FT$ — база ШПС; E_6/N_0 — реальное значение отношения сигнал/шум в системе.

Напомним, что в системах с частотным разделением пропускная способность $M = F/R$.

Из выражения (7.22) видно, что при асинхронной работе с ростом базы B , а следовательно, и занимаемой сигналами ШПС полосы, при прочих равных условиях, допустимое число одновременно действующих каналов увеличивается.

Оценим, как соотносятся пропускные способности асинхронных систем с ШПС и с частотным разделением. Для упрощения изложения будем полагать, что свойственные частотному разделению защитные интервалы отсутствуют. В этих условиях, учитывая, что пропускная способность систем с частотным разделением $M_{\text{чр}} = F/R$, получим

$$\frac{M}{M_{\text{чр}}} \approx \frac{N_{\Sigma}}{E_6} - \frac{N_0}{E_6} < 1. \quad (7.23)$$

С ростом энергетического запаса E_6/N_0 относительная пропускная способность асинхронных систем с ШПС растет, но даже при $E_6/N_0 \rightarrow \infty$, как следует из (7.23), она остается в E_6/N_{Σ} раз меньше, чем в системах с частотным разделением.

Рассмотрим за счет каких действий допустимое число одновременно действующих каналов при использовании ШПС можно увеличить.

Известно, что в подвижных системах связи основным видом предоставляемых услуг является телефонная связь, а основным источником информации — речь. Также хорошо из-

вестно, что в занятом телефонном канале передача речевых сигналов занимает не более 1/3 времени и примерно 2/3 времени приходится на паузы. При выключении или значительном снижении мощности излучения во время пауз в речи при телефонном разговоре мощность взаимных помех будет снижаться пропорционально времени пауз. За счет этого допустимое число одновременно действующих каналов, а следовательно, и пропускная способность системы может быть в два-три раза увеличены. Дополнительной полосы в данном случае не требуется. Также в два-три раза увеличивается и эффективность использования спектра, занимаемого системой с ШПС.

В системах связи с простыми сигналами и частотным разделением за счет этого повышается эффективность использования мощности передатчика ретранслятора или базовой станции. Однако эффективность использования полосы не повышается, так как выделенная каждому абоненту на время сеанса полоса частот сохраняется за ним даже в паузах речи.

С учетом этого пропускная способность системы с ШПС может быть определена следующим выражением:

$$M = \frac{1}{\alpha} \frac{F}{R} \left(\frac{N_{\Sigma}}{E_6} - \frac{N_0}{E_6} \right) + 1, \quad (7.24)$$

где $\alpha \approx (1/2 - 1/3)$ — коэффициент активности абонента.

Другая особенность систем радиосвязи с ШПС связана с использованием в ретрансляторе или на базовой станции узконаправленных многолучевых антенн. Известно, что применение таких антенн позволяет упростить требования к наиболее массовым земным станциям, уменьшить ограничения на пропускную способность системы по полосе за счет повторного использования частот в разных лучах. Однако из-за «неидеальности» диаграмм направленности антенн в системе связи с L лучами и простыми сигналами одну и ту же частоту можно использовать не чаще чем $L/3$ или $L/7$ раз, а в соседних лучах должны использоваться разные полосы частот. В случае ШПС разнос частот сигналов в разных лучах не требуется, а мощность помехи от сигналов из соседних лучей, даже «незначительно» подавленная за счет многолучевых бортовых антенн с «неидеальными» диаграммами направленности, оказывается значительно меньше мощности взаимных помех от мешающих сигналов своего луча. Далее помехи будут подавлены за счет кодового разделения ШПС. Таким образом, эффективность использования спектра в системе с ШПС может значительно превысить эффективность применения спектра в системе с частотным разделением.

Дополнительное увеличение эффективности использования спектра примерно на 60% достигается за счет возможного разделения сигналов по поляризации. Известно [7.10], что в подвижных спутниковых станциях трудно подавить сигнал с нежелательной поляризацией более чем на 6 дБ. Этого совершенно не достаточно для разделения сигналов в системах с частотным или временным разделением. Для систем с ШПС и такое подавление приводит к заметному снижению взаимных помех, что также повышает эффективность использования спектра, хотя следует отметить, что широкого применения на практике этот метод пока не нашел.

В мобильных системах связи с простыми сигналами эффективность использования спектра снижается за счет того, что при большой неопределенности по доплеровским смещениям частоты и изменениях времени прихода принимаемых сигналов для снижения взаимных помех приходится вводить защитные интервалы по частоте или по времени. В системах с ШПС защитные интервалы не требуются.

В приведенных выше положениях предполагалось, что все входящие на ретранслятор сигналы имеют одинаковую мощность. В случае разных мощностей пропускная способность системы резко снижается. Разброс мощностей на 3 дБ приводит к уменьшению

эффективности примерно в два раза. Поэтому регулировка мощностей абонентских станций с целью выравнивания мощностей сигналов, приходящих на ретранслятор, в системе с ШПС обязательна.

Как уже отмечалось выше, пропускная способность в широкополосных системах в основном ограничивается уровнем системных помех и появление каждого «нового» активного абонента приводит к его увеличению. Важной особенностью таких систем является то, что возрастание уровня системных помех (мягкая перегрузка) приводит не к отказу в предоставлении услуги, а лишь к временному снижению ее качества, что позволяет повысить среднюю загрузку системы и, следовательно, эффективность использования спектра.

В системах связи с простыми сигналами эффективность использования спектра дополнительно снижается за счет того, что при большой неопределенности частоты, вызванной эффектом Доплера, и изменениях времени прихода принимаемых сигналов для снижения взаимных помех приходится вводить защитные интервалы по частоте. В системах с ШПС защитные интервалы не требуются. В результате эффективность использования спектра в системе с ШПС оказывается значительно выше, чем в системах с простыми сигналами.

7.3.2. Защита от замираний

В системах спутниковой связи с малыми станциями имеют место замирания, обусловленные многолучевым распространением радиоволн. Малые антенны земных станций имеют широкие диаграммы направленности и поэтому не могут разделить прямой и отраженные лучи. В случае ШПС, принимаемые по различным лучам сигналы, могут быть разделены, а результирующий сигнал не подвержен замираниям, вызываемым многолучевым распространением. При приеме сигналы разделенных лучей могут выделяться и когерентно складываться. Эксперименты [7.10] показали, что в реальных условиях энергетический выигрыш от реализации ШПС при многолучевом распространении составил 6–9 дБ (более подробно см. гл. 9).

7.3.3. Электромагнитная совместимость

Достоинством систем связи с ШПС признается хорошая электромагнитная совместимость с существующими радиосредствами. Так, применение станций с малыми антеннами предполагает увеличение эквивалентной изотропно-излучаемой мощности (ЭИИМ) ретранслятора. Расширение спектра позволяет соблюсти установленные нормы на спектральную плотность потока излучаемой мощности. Станции с ШПС могут работать на вторичной основе в общем частотном диапазоне с существующими узкополосными средствами.

7.3.4. Эффективность использования мощности бортового передатчика

Применение ШПС открывает возможность построения эффективных спутниковых систем связи с прямой ретрансляцией сигналов абонентских станций. В системах с простыми сигналами при частотном разделении каналов передатчик ретранслятора должен находиться в линейном режиме, в противном случае нелинейные продукты третьего порядка могут поразить отдельные частотные каналы. При этом средняя мощность передатчика на 3–6 дБ ниже мощности насыщения. При использовании ШПС и выравнивании уровней сигналов на входе ретранслятора переход в нелинейный режим передатчика также приводит к образованию нелинейных продуктов, но они, как правило, не представляют опасности для систем с кодовым разделением, поэтому пропускная способность системы с ШПС может быть увеличена в два-три раза за счет перевода передатчика ретранслятора в режим, близкий к нелинейному [7.10].

7.4. Псевдослучайная перестройка рабочей частоты

В широкополосных системах связи с ППРЧ вся полоса частот, выделенная для работы системы, делится на относительно большое число прилегающих друг к другу элементарных полос (частотных позиций). В произвольный «момент» времени спектр передаваемого некоторой станцией сигнала занимает одну или (реже) несколько элементарных рабочих (частотных) полос. Выбор элементарных полос осуществляется по некоторому псевдослучайному закону, формируемому генератором ПСП (см. рис. 7.2).

В широкополосной системе со многими абонентами, например, в спутниковых или сотовых системах с частотным доступом (МДЧР) «прыжки» по частоте могут осуществляться синхронно с сохранением исходного частотного разделения (рис. 7.6). Имеется две разновидности ППРЧ: быстрая и медленная. При *быстрой* ППРЧ на длительности каждого информационного или кодированного символа происходит несколько «прыжков» по частоте. Однако метод быстрой ППРЧ в системах связи широкого применения не получил. Особенности его реализации мало отличаются от рассмотренного выше метода ШПС.

При *медленной* ППРЧ на длительности каждой частотной позиции передается несколько информационных или кодированных символов. Этот метод достаточно широко применяется для борьбы с замираниями (разнесение по частоте). Кроме того, метод медленной ППРЧ используется в сверхширокополосных системах связи при необходимости работать в условиях организованного радиопротиводействия. Во многом это объясняется относительно простой реализацией аппаратуры формирования и обработки сверхширокополосных сигналов и тем, что при медленной ППРЧ намного проще, чем при ШПС решаются проблемы синхронизации. (Так, требуемая точность синхронизации при одинаковых полосах передачи при ШПС в базу раз выше, чем для медленной ППРЧ.) Степень расширения спектра при ППРЧ (база сигналов) определяется отношением полосы перестройки частоты к информационной полосе.

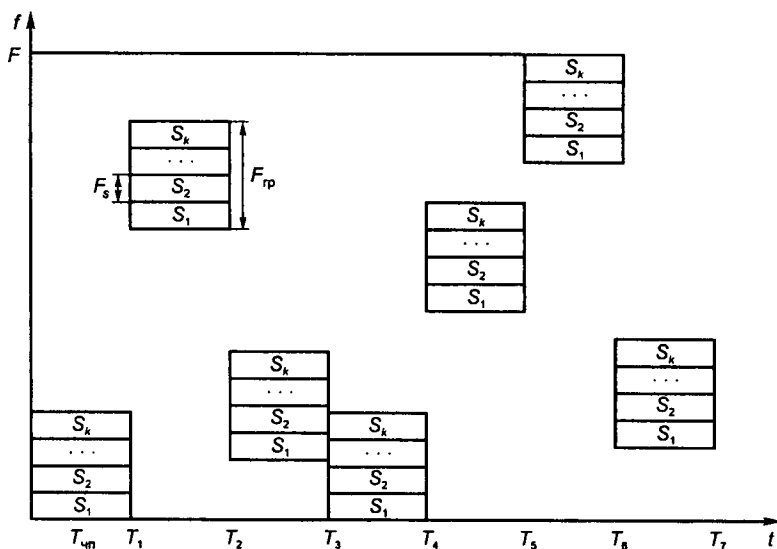


Рис. 7.6. Частотно-временное представление k сигналов с ППРЧ

Следует отметить некоторые особенности помехоустойчивого кодирования при медленной ППРЧ. Как уже отмечалось, на длительности каждой частотной позиции передается несколько информационных, или кодированных символов. Предположим для простоты, что на длительности каждой частотной позиции, например, при блоковом кодировании передается одно или несколько кодовых слов. Тогда, если эта частотная позиция окажется подверженной воздействию преднамеренных помех или замираний, то имеет место группирование символьных ошибок в пакеты, и с высокой вероятностью эти одно или несколько кодовых слов окажутся принятыми с ошибками. Эффект от такого кодирования окажется нулевым. Для того чтобы избежать этого, для каждого кодового слова необходимо обеспечить декорреляцию ошибок до декодирования. Достаточно провести перемежение символов, т.е. кодированную информацию расположить таким образом, чтобы на длительности каждой частотной позиции передавались только символы различных кодовых слов.

При приеме перед декодированием необходимо провести деперемежение символов так, чтобы обеспечивалась декорреляция ошибок в символах каждого кодового слова. Это восстанавливает эффективность кодирования. Частным случаем такого кодирования может быть простое повторение передачи одинаковых блоков информации на разных рабочих частотах с последующим мажоритарным декодированием. Более подробно оценка помехоустойчивости или помехозащищенности будет рассматриваться ниже.

7.5. Помехозащищенность широкополосных систем связи

К числу наиболее важных параметров, характеризующих системы передачи информации, в том числе и системы связи, относятся помехоустойчивость и помехозащищенность. Помехоустойчивость и помехозащищенность это способность систем связи обеспечить требуемое качество связи при воздействии помех. В чем же их отличие?

Под *помехоустойчивостью* (см. гл. 3) понимают способность приемника обеспечить в процессе демодуляции требуемое качество связи при воздействии помех естественной природы типа аддитивного БГШ. Количественно помехоустойчивость определяется величиной отношения E_b/N_0 , которое необходимо для обеспечения заданного качества связи, например заданной вероятности ошибки в приеме бита. Та система будет более помехоустойчивой, которая обеспечивает заданное качество связи при меньшем значении E_b/N_0 при прочих равных условиях.

Вопросы оценки потенциальной помехоустойчивости радиосистем рассмотрены ранее в гл. 3. Здесь отметим только один фундаментальный результат, полученный основателем теории потенциальной помехоустойчивости В.А. Котельниковым: *при воздействии помех типа аддитивного БГШ помехоустойчивость оптимального приемника зависит от отношения энергии сигнала E_b к спектральной плотности шума N_0 , коэффициентов корреляции между сигналами и не зависит от формы сигналов.* Следовательно, помехоустойчивость приема простых сигналов ($B \approx 1$) и широкополосных (и ШПС, и ППРЧ) при прочих равных условиях всегда одинакова.

Под *помехозащищенностью* понимают способность систем связи обеспечить требуемое качество связи при воздействии преднамеренных помех, т.е. помех специально создаваемых средствами радиопротиводействия. Помехозащищенность количественно определяется как отношение мощности полезного сигнала к мощности преднамеренных помех в полосе полезного сигнала $P_c/P_{\text{пом}}$, при котором еще обеспечивается требуемое качество связи. Качество связи также характеризуется вероятностью ошибки в приеме бита.

В современных системах связи специального назначения для обеспечения эффективной защиты от преднамеренных и случайных помех совместно используются методы, основанные на разделении по направлению прихода полезных сигналов и преднамеренных помех и методы широкополосной передачи. Оба метода в настоящее время активно разрабатываются. Поскольку каждый из них почти независим от другого, ниже особое внимание будет уделено только анализу помехозащищенности широкополосных систем связи. Рассмотрение методов, основанных на разделении по направлению прихода полезных сигналов и преднамеренных помех, выходит за рамки настоящей работы.

Анализ помехозащищенности при использовании ШПС рассмотрен и достаточно подробно в ряде работ (см., например, [7.4, 7.11]). Однако анализ помехозащищенности при использовании ППРЧ в качестве метода расширения спектра для защиты от преднамеренных помех имеет свои особенности. Так, мало эффективные преднамеренные помехи для ШПС оказываются очень эффективными для ППРЧ, и наоборот.

7.5.1. Анализ помехозащищенности при ППРЧ

Определим, какие помехи, например, в классе флуктуационных гауссовских в условиях организованного радиопротиводействия со стороны возможного постановщика помех оказываются наиболее эффективными при ППРЧ.

Стратегия постановщика помех базируется на реализации такого распределения помехи по полосе частот, занимаемой спектром сигналов с ППРЧ, при котором при фиксированной средней мощности помехи J_{cp} достигается максимум вероятности ошибки.

Известно [7.12], что наиболее эффективным признается такое распределение, когда помеха поражает только часть ρF общей полосы частот F , занимаемой спектром сигналов с ППРЧ. При этом спектральная плотность помехи

$$N_{\Sigma} = \begin{cases} N_n(\rho) + N_0, & \text{с вероятностью } \rho; \\ N_0, & \text{с вероятностью } 1 - \rho, \end{cases} \quad (7.25)$$

где $N_n(\rho) = J/(\rho F)$ — спектральная плотность преднамеренной помехи; J — мощность преднамеренной помехи; ρ — часть всей рабочей полосы, пораженная преднамеренной помехой ($0 \leq \rho \leq 1$).

Помехозащищенность ППРЧ при ОФМ. Как уже отмечалось, свойственные ППРЧ псевдослучайные скачки частоты влекут за собой и случайные скачки фазы принимаемых сигналов. Следствием этого также является необходимость реализации некогерентного приема, т.е. приема сигналов со случайной начальной фазой. Сначала рассмотрим наиболее простой случай двоичной передачи с модуляцией методом относительной фазовой манипуляции (ОФМ).

Известно, что при помехе типа аддитивного БГШ оптимальный некогерентный прием сигналов с ОФМ при ППРЧ реализуется технически достаточно просто. Вероятность ошибки на бит в этом случае описывается выражением (3.39)

$$P_{\text{ошб}} = \frac{1}{2} \exp(-E_6/N_0). \quad (7.26)$$

При помехе в виде суммы БГШ и преднамеренной помехи, поражающей только часть всей рабочей полосы частот, используя (7.25), получим

$$P_{\text{ошб}} = \rho \frac{1}{2} \exp\left(-\rho \frac{E_6}{(J/F) + \rho N_0}\right) + (1 - \rho) \frac{1}{2} \exp\left(-\frac{E_6}{N_0}\right). \quad (7.27)$$

Как правило, при приеме сигналов на фоне преднамеренных помех спектральная плотность помехи $J/(\rho F) \gg N_0$, тогда

$$P_{\text{ошб}} \approx \rho \frac{1}{2} \exp\left(-\rho \frac{E_6}{(J/F)}\right). \quad (7.28)$$

Легко показать [7.13], что оптимальное значение ρ , приводящее к максимуму $P_{\text{ошб}}$, при фиксированной мощности помехи J определяется выражением

$$\rho_{\text{опт}} = \frac{J/F}{E_6} \leq 1. \quad (7.29)$$

При этом имеем

$$\max_{\rho} P_{\text{ошб}} = \frac{1}{2} \frac{e^{-1}}{\frac{E_6}{J/F}}. \quad (7.30)$$

Отметим, что исходная экспоненциальная зависимость $P_{\text{ошб}}$ от E_6/N_n (см. (7.26)) при помехах, поражающих оптимальную часть полосы частот, превращается в обратно пропорциональную (7.30). На рис. 7.7 приведена зависимость вероятности ошибки на бит от $E_6/(J/F)$ при помехах типа БГШ и оптимальных прерывистых по частоте помехах. Как видно из рисунка, при ОФМ эффективность преднамеренной помехи, поражающей оптимальную часть полосы частот, значительно превышает эффективность помехи той же мощности, поражающей всю полосу частот, занимаемую спектром сигналов с ППРЧ.

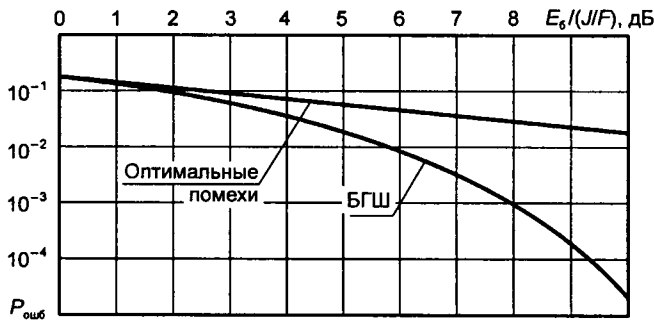


Рис. 7.7. Помехозащищенность ППРЧ относительно помех типа БГШ и оптимальных помех

Помехозащищенность ППРЧ при m -ичной манипуляции. Как уже отмечалось в 4.3, при некогерентном приеме ($m > 2$) наилучшую помехоустойчивость обеспечивает использование ортогональных сигналов. Ниже в качестве ортогональных сигналов будем рассматривать либо m -ЧМ (m -ичная частотная манипуляция), либо m -ШПС. Для дальнейшего анализа сделаем следующие ограничения. Пусть полоса преднамеренной помехи ρF всегда больше полосы, занимаемой суммарным спектром m -ичных сигналов (без учета расширения спектра из-за ППРЧ) и на длительности частотной позиции располагается только один m -ичный сигнал. Тогда, используя выражение (3.38), при помехах, поражающих часть рабочей полосы частот, получим

$$P_{\text{ошб}} = \frac{1}{2(m-1)} \rho \sum_{i=2}^m (-1)^i C_m^i \exp\left(-\frac{\rho E_6 k \frac{i-1}{i}}{(J/F)}\right) \leq 2^{t-2} \rho \exp\left(-\rho \frac{E_6 k}{(2J/F)}\right), \quad (7.31)$$

где $k = \log_2 m$. Отметим, что первое слагаемое (при $i = 2$) в выражении (7.31) представляет собой верхнюю оценку для вероятности ошибки, поскольку каждое последующее слагаемое меньше предыдущего, а их знаки чередуются, т.е.

$$P_{\text{ошб}} < 2^{t-2} \rho \exp\left(-\rho \frac{E_6 k}{(2J/F)}\right). \quad (7.32)$$

Для нахождения оптимального значения ρ , приводящего $P_{\text{ошб}}$ к максимуму, продифференцируем правую часть выражения (7.31) по ρ :

$$\begin{aligned} \frac{dP_{\text{ошб}}}{d\rho} = \frac{m}{2(m-1)} \left\{ \sum_{i=2}^m (-1)^i C_m^i \exp\left(-\frac{\rho E_6 k \frac{i-1}{i}}{(J/F)}\right) - \right. \\ \left. - \rho \sum_{i=2}^m (-1)^i C_m^i \frac{E_6 k \frac{i-1}{i}}{(J/F)} \exp\left(-\frac{\rho E_6 k \frac{i-1}{i}}{(J/F)}\right) \right\}. \end{aligned} \quad (7.33)$$

Приравняв (7.33) нулю, получим

$$\begin{aligned} \frac{\rho_{\text{опт}} \frac{E_6 k}{(J/F)} - 1}{\rho_{\text{опт}} \frac{E_6 k}{(J/F)}} = \frac{\sum_{i=2}^m (-1)^i C_m^i \frac{1}{i} \exp\left(-\frac{\rho_{\text{опт}} E_6 k \frac{1}{i}}{(J/F)}\right)}{\sum_{i=2}^m (-1)^i C_m^i \exp\left(-\frac{\rho_{\text{опт}} E_6 k \frac{1}{i}}{(J/F)}\right)}. \end{aligned} \quad (7.34)$$

В общем случае при $m > 2$ решение уравнения (7.34) удастся найти только численными методами. Для $m = 2$ уравнение (7.34) решается аналитически. В результате имеем

$$\rho_{\text{опт}} = \frac{(2J/F)}{E_6} \leq 1; \quad \max_{\rho} P_{\text{ошб}} = \frac{1}{e \frac{E_6}{(J/F)}}. \quad (7.35)$$

Для значений m обычно используемых на практике результаты решения уравнения (7.35) приведены в табл. 7.2.

Таблица 7.2. Значения верхней оценки вероятности ошибки для различных m

№ п/п	m	$\rho_{\text{опт}}(m)$	$\max_{\rho} P_{\text{ошб}}(m)$	Верхняя оценка $\max_{\rho} P_{\text{ошб}}(m)$
1	2	3	4	5
1	2	2А	0,36А	0,36А
2	4	1,17А	0,23А	0,36А
3	8	0,93А	0,19А	0,49А
4	16	0,87А	0,18А	0,73А
5	32	0,80А	0,17А	1,17А

В этой же таблице в графе 4 приведены соответствующие выражения для $\max_{\rho} P_{\text{ошб}}(m)$, полученные подстановкой $\rho_{\text{опт}}(m)$ в (7.31).

Верхняя оценка вероятности ошибки при действии помех, поражающих оптимальную часть полосы частот, находится подстановкой $\rho_{\text{опт}}(m)$ в правую часть выражения (7.31). В результате получим

$$\max_{\rho} P_{\text{ошб}}(m) \leq \frac{m}{2k} \frac{1}{e^{\frac{E_6}{J/F}}}. \quad (7.36)$$

Вычисленные значения верхней оценки вероятности ошибки для различных m приведены в графе 5. В табл. 7.2 $A = \frac{(J/F)}{E_6}$ — величина обратно пропорциональная отношению энергии бита к спектральной плотности преднамеренной помехи. При получении этого результата никаких ограничений на выбор типа сигналов (m -ЧМ или m -ШПС) не делалось.

На рис. 7.8 приведены зависимости вероятности ошибки на бит от $\frac{E_6}{(J/F)}$ для $m = 2, 4, 8, 16$ и 32 .

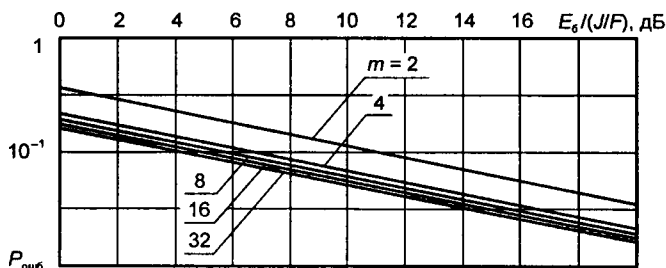


Рис. 7.8. Зависимости вероятности ошибки от $\frac{E_6}{(J/F)}$

Из результатов расчетов следует, что как и для ОФМ при помехах, поражающих оптимальную часть полосы рабочих частот, помехозащищенность приема m -ичных ортогональных сигналов чрезвычайно низка. С ростом m вероятность ошибки на бит от $\frac{E_6}{(J/F)}$

незначительно убывает. В случае, когда спектр помехи на пораженной частотной позиции занимает полосу порядка $1/T_m$, где T_m — длительность m -ичного символа, т.е. помеха относительно ширины группового спектра сигналов m -ЧМ или m -ШПС окажется узкополосной, помехозащищенность m -ЧМ или m -ШПС будет различной, хотя и останется чрезвычайно низкой.

Пусть постановщик преднамеренных помех размещает узкополосные (например, гармонические) составляющие помехи так, чтобы только одна из них попадала на участок полосы, занимаемый групповым спектром m -ичных ортогональных сигналов, поражая при этом l частотных позиций. Число независимых частотных позиций сигналов с ППРЧ

$$N = \frac{F}{R_6} \frac{k}{m}, \quad (7.37)$$

где R_6 — скорость передачи двоичной информации.

Вероятность ошибки при m -ЧМ и такой помехе с гармоническими составляющими

$$P_{\text{ошб}}(m) = \frac{m}{2(m-1)} \frac{l}{N} \frac{(m-1)}{m}. \quad (7.38)$$

Выражение (7.38) справедливо только при $(J/l) > P_c$, где P_c — мощность полезного сигнала. Подставляя (7.37) в (7.38), после несложных преобразований получим

$$P_{\text{ошб}}(m) = \frac{m}{2k \frac{E_6}{(J/F)}}. \quad (7.39)$$

При m -ШПС узкополосная помеха произвольной структуры, поразившая частотную позицию, занятую спектром полезного сигнала, одновременно попадает и во все m приемных каналов. Полагая, что распределение узкополосной помехи на выходе приемника ШПС (согласованного фильтра) на основании центральной предельной теоремы нормализуется, получим выражения для вероятности ошибки, совпадающие с выражениями графы 4 в табл. 7.2. Многочисленные экспериментальные исследования, проведенные рядом авторов, также показывают, что реальная помехозащищенность приема m -ЧМ относительно узкополосных помех примерно на (3–4) дБ, т.е. примерно в e раз ниже, чем при флукуационных.

7.5.2. Эффективность и особенности кодирования при ППРЧ

Эффективным средством, обеспечивающим резкое улучшение помехозащищенности при помехах, поражающих оптимальную часть полосы частот, занимаемую спектрами сигналов с ППРЧ, является избыточное кодирование с разнесением (повторением передачи) символов кодовых слов по различным частотным позициям. Как достаточно простой случай такого кодирования рассмотрим кодирование повторением передачи каждого символа m -ичного кода L раз на различных частотных позициях и мягким декодированием при некогерентном приеме. Анализ помехозащищенности проведем, используя в качестве оценки помехозащищенности границу Чернова [7.12]

$$P_{\text{ошб}}(m) < \frac{m}{4} \left\{ \frac{\rho}{1-\lambda^2} \exp \left[-\frac{\lambda}{1+\lambda} \left(\rho \frac{E_6 k}{(LJ/F)} \right) \right] \right\}^L = \hat{P}_{\text{ошб}}(m), \quad (7.40)$$

где λ — параметр Чернова.

В выражении (7.40) имеются три параметра, по которым следует провести поиск экстремума: λ , ρ и L .

Аналогично предыдущему выводу найдем $\max_{\rho} P_{\text{ошб}}(m)$:

$$\frac{d\hat{P}_{\text{ошб}}(m)}{d\rho} = \frac{L}{A} \rho^{L-1} \exp(-BCL) - \frac{\rho L}{A} \exp(-BCL\rho) = 0, \quad (7.41)$$

где $A = \frac{1}{1-\lambda^2}$, $B = \frac{\lambda}{1+\lambda}$, $C = \frac{E_6 k}{(LJ/F)}$.

Из уравнения (7.41) находим

$$\rho_{\text{опт}} = \frac{1}{BC} = \frac{1+\lambda}{\lambda} \frac{(LJ/F)}{E_6 k}. \quad (7.42)$$

Подставляя (7.42) в (7.41), получим

$$\max_{\rho} \hat{P}_{\text{ошб}}(m) \leq \frac{m}{4} \left[\left(\frac{1+\lambda}{\lambda} \right) \frac{(LJ/F)}{E_6 k} \left(\frac{1}{1-\lambda^2} \right) e^{-1} \right]^L. \quad (7.43)$$

С учетом (7.42) оптимизируем (7.41) по параметру Чернова λ :

$$\frac{d\hat{P}_{\text{ошб}}(m)}{d\lambda} = 0. \quad (7.44)$$

В результате получим, что $\max_{\lambda} \hat{P}_{\text{ошб}}(m)$ имеет место при

$$\lambda_{\text{опт}} = 1/2. \quad (7.45)$$

С учетом (7.45) из выражения (7.43) получим

$$\min_{\lambda} \max_{\rho} \hat{P}_{\text{ошб}}(m) \leq m \left[\frac{(LJ/F)}{E_6 k} e^{-1} \right]^L, \quad (7.46)$$

$$\rho_{\text{опт}} = \frac{(3LJ/F)}{E_6 k}. \quad (7.47)$$

И последнее, оптимизируем (7.40) по L с учетом (7.45) и (7.47).

Пусть $\frac{d\hat{P}_{\text{ошб}}(m)}{dL} = 0$, тогда $L_{\text{опт}} = \frac{E_6 k}{(4J/F)}$.

Таким образом, при оптимальном разнесении m -ичных символов по различным частотным позициям верхняя граница вероятности ошибки при помехах, поражающих оптимальную часть рабочей полосы частот сигналов с ППРЧ, имеет вид

$$\hat{P}_{\text{ошб}}(m) = \frac{m}{4} \exp\left(-\frac{k}{4} \frac{E_6}{(J/F)}\right). \quad (7.48)$$

Из выражения (7.48) следует, что оптимальное разнесение (повторение) передачи каждого символа m -ичного кода на L различных частотных позиций возвращает вероятности ошибки (при оптимизации по верхней границе) первоначальную экспоненциальную зависимость

от отношения энергии сигнала к спектральной плотности помехи $\left(\frac{E_6}{(J/F)}\right)$.

Для $m = 2, 8$ и 32 эти зависимости приведены на рис. 7.9.

Для $m = 2$ ($k = 1$)

$$\hat{P}_{\text{ошб}}(2) = \frac{1}{2} \exp\left(-\frac{E_6}{(4J/F)}\right), \quad (7.49)$$

т.е. верхняя граница $\hat{P}_{\text{ошб}}(2)$ при помехах, поражающих оптимальную часть рабочей полосы сигналов с ППРЧ, уступает 3 дБ точному значению вероятности ошибки $P_{\text{ошб}}(2)$ относительно помех типа БГШ.

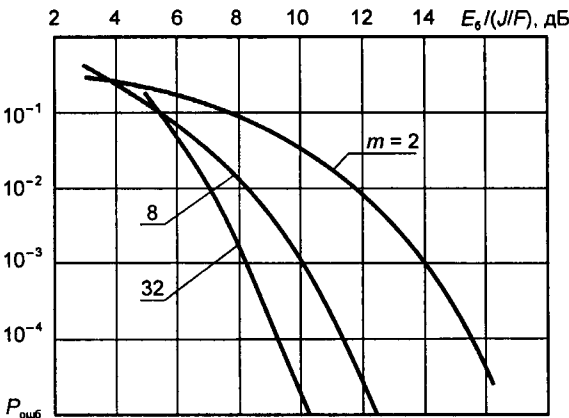


Рис. 7.9. Зависимости вероятности ошибки при оптимальном разнесении

При $m = 2$ более просто на практике реализуется алгоритм жесткого декодирования, когда сначала выносится первое решение о значении кодированного символа, передаваемого на L частотных позициях, а затем второе с использованием мажоритарного правила принятия решений. Вероятность ошибки относительно помех, поражающих оптимальную часть рабочей полосы частот сигналов с ППРЧ, определяется известным выражением (см. (4.42))

$$P_{\text{ошб}}(2) = \sum_{i=\frac{L+1}{2}}^L C_L^i p^i (1-p)^{L-i}, \tag{7.50}$$

где p определяется выражением (7.36), в котором E_6 заменено на E_c , а энергия символа $E_c = E_6/L$. На рис. 7.10 приведены зависимости вероятности ошибки на бит от $\frac{E_6}{(J/F)}$ для $m = 2$ и $L = 1, 3, 5$ и 7 повторений-разнесений.

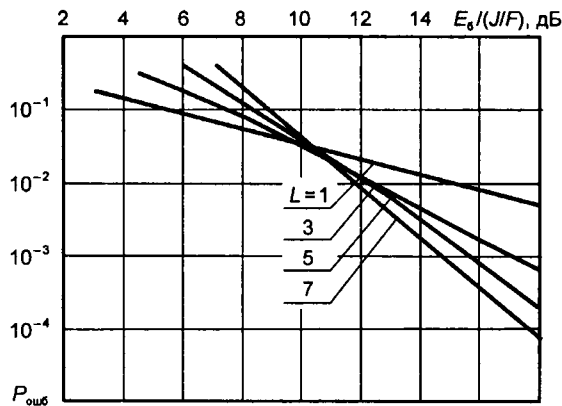


Рис. 7.10. Зависимости вероятности ошибки на бит

При использовании избыточного кодирования (n, k) кодом с кодовым расстоянием d в сочетании с передачей символов каждого кодового слова на различных частотных позициях вероятность ошибки относительно помех, поражающих оптимальную часть рабочей полосы сигналов с ППРЧ, определяется выражением

$$P_{\text{ошб}}(2) \leq \frac{1}{n} \sum_{i=\frac{d+1}{2}}^L i C_L^i p^i (1-p)^{L-i}, \tag{7.51}$$

где p также определяется выражением (7.36), в котором E_6 заменено на E_c , а $E_c = E_6(k/n)$ — энергия кодированного символа.

Эффективность кодирования относительно помех, поражающих оптимальную часть рабочей полосы сигналов с ППРЧ, может быть еще более усилена, если при декодировании обнаруживать пораженные преднамеренной помехой частотные позиции. Оригинальный алгоритм обнаружения пораженных частотных позиций предложен в работе [7.14], однако его изложение выходит за рамки данной работы.

7.5.3. Анализ помехозащищенности при ШПС

Определим, какие помехи в условиях организованного радиопротиводействия со стороны возможного постановщика помех оказываются наиболее эффективными при использовании ШПС. Здесь стратегия постановщика помех основывается на реализации такого распределения помехи по времени, при котором при фиксированной средней мощности помехи $J_{п.ср}$ достигается максимум вероятности ошибки [7.13]. Как и при ППРЧ будем рассматривать случай некогерентного приема. Пусть такая помеха поражает s бит сигнала из S , т.е. вероятность поражения бита помехой $\rho = s/S$. Тогда при использовании ШПС оптимальное значение ρ , приводящее к $\max_{\rho} \hat{P}_{ошб}(m)$, и само максимальное значение $\max_{\rho} \hat{P}_{ошб}(m)$, как и в слу-

чае ППРЧ без кодирования, определяются выражениями (7.35). При кодировании и оптимальном разнесении-перемежении, так чтобы ошибки кодированных символов после перемежения не группировались в пачки, значение вероятности ошибки при ШПС также будет удовлетворять, как и при ППРЧ, оценкам (7.48) и (7.51). Однако, несмотря на то, что выражения для максимума верхней оценки вероятности ошибки и при ШПС, и при ППРЧ совпадают, имеются и определенные отличия. Так, передатчик прерывистых по времени помех должен иметь «импульсную» мощность в $1/\rho$ раз выше средней. При ППРЧ средняя и «импульсная» мощности передатчика помех могут быть одинаковыми. (Здесь под «импульсной» мощностью понимается мощность помехи только в интервале ее излучения.)

Таким образом, помехозащищенность и при ППРЧ, и при ШПС относительно помех, оптимальных для каждого вида широкополосной передачи, одинакова и без кодирования очень низка. Вероятность ошибки здесь обратно пропорциональна отношению энергии сигнала к спектральной плотности преднамеренных помех. Кодирование практически возвращает вероятности ошибки исходную экспоненциальную зависимость от отношения энергии сигнала к спектральной плотности преднамеренных помех.

Глава 8

ВЕРОЯТНОСТНЫЕ МОДЕЛИ МНОГОЛУЧЕВЫХ РАДИОКАНАЛОВ

8.1. Многолучевой радиоканал

Распространение радиоволн накладывает фундаментальные ограничения на качество передачи информации по радиоканалам. Путь распространения радиоволны от передатчика к приемнику может иметь самые разнообразные геометрические представления. В простейшем случае прямой видимости в свободном пространстве это прямая линия. При наличии одного отражателя, не прерывающего линию прямой видимости, приемник принимает электромагнитные волны, распространяющиеся по двум путям — по линии прямой видимости (прямой луч) и ломаной линии «передатчик–отражатель–приемник» (отраженный луч с углом прихода на приемную антенну α). В реальных условиях в точку приема может приходиться много лучей, прошедших от передатчика к приемнику по очень сложным путям, имеющим разную длину. В условиях интенсивной городской застройки число возможных путей прихода радиоволн к приемнику может оказаться неограниченно большим, когда значения углов их прихода заполняют непрерывный интервал $[0, 2\pi]$. Таким образом, отражение радиоволн от различных препятствий на пути их распространения является первым существенным эффектом, который приходится учитывать при построении вероятностных моделей различных радиоканалов.

Во многих современных радиосистемах передачи информации типичным условием функционирования считается отсутствие прямой видимости между антеннами передатчика и приемника. Более того, большая часть энергии принимаемого суммарного электромагнитного поля обусловлена эффектом дифракции — рассеянием распространяющейся радиоволны на препятствиях в виде зданий, неровностей рельефа местности и других объектов. Таковыми являются, например, условия организации связи с подвижными объектами, когда передатчик и приемник системы передачи располагаются в интенсивной городской застройке. Передатчик и приемник могут находиться внутри зданий, в лесных массивах и т.д., когда приходится учитывать эффекты проникновения электромагнитных волн в препятствия, их рассеяние на различных предметах, находящихся в зоне их распространения.

Цель данного раздела — привести описание сигнала как функции времени в точке приема при наличии многих путей распространения радиоволн от передатчика к приемнику. Полагаем, что форма сигнала $s(t)$, излучаемого передатчиком, известна, а радиоканал вдоль каждого пути распространения радиоволны не вносит искажений формы сигнала, т.е. является широкополосным и имеет постоянные параметры. В этом случае сигнал в приемнике, созданный радиоволной, пришедшей к антенне приемника по пути длиной r , можно записать следующим образом:

$$s_{\text{пр}}(t) = ks(t - \tau), \quad (8.1)$$

где $\tau = r/c$ — время распространения радиоволны со скоростью c от антенны передатчика до антенны приемника; k — коэффициент, учитывающий изменение мощности сигнала при распространении волны вдоль рассматриваемого пути. Средняя мощность принимаемого сигнала на любом интервале времени длительностью T определится обычным образом:

$$P_{\text{пр}} = \frac{1}{T} \int_{t_0}^{t_0+T} k^2 s^2(t - \tau) dt = k^2 \frac{1}{T} \int_{t_0}^{t_0+T} s^2(t - \tau) dt = k^2, \quad (8.2)$$

если принять равной единице среднюю мощность сигнала, излучаемого передатчиком.

Если сигнал в приемнике создается n радиоволнами, каждая из которых прошла по своему пути распространения длиной r_i , $i = 1, 2, \dots, n$, то вместо (8.1) можно записать

$$s_{\text{пр}}(t) = \sum_{i=1}^n k_i s(t - \tau_i). \quad (8.3)$$

Очевидно, что вместо (8.3) можно записать более общее выражение

$$s_{\text{пр}}(t) = \int_{-\infty}^t h(\tau) s(t - \tau) d\tau, \quad (8.4)$$

если принять, что канал является линейным устройством с импульсной характеристикой

$$h(\tau) = \sum_{i=1}^n k_i \delta(\tau - \tau_i), \quad (8.5)$$

где $\delta(\tau)$ — дельта-функция Дирака. Понятие импульсной характеристики канала передачи широко используется в настоящее время при описании любых каналов связи. Однако для

радиоканала здесь требуется одно существенное уточнение, которое можно пояснить с помощью рис. 8.1.

Здесь антенны передатчика П и приемника Пр изображены в плане и помещены в фокусы эллипса с большой $2a$ и малой $2b$ осями. На рисунке изображен путь прямой волны длиной r_0 и пути двух отраженных от препятствий волн с длиной $r_1 = r'_1 + r''_1$ и $r_2 = r'_2 + r''_2$. Если точки отражения волн находятся на эллипсе, то $r_1 = r_2 = 2a$, так что время распро-

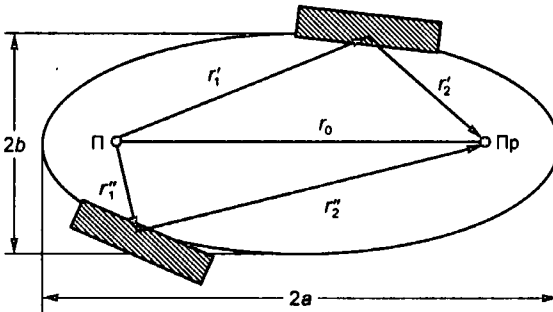


Рис. 8.1. Формирование одного луча в многолучевом канале

странения радиоволн от антенны передатчика до антенны приемника оказывается одинаковым, т.е. $\tau_1 = \tau_2 = \tau_a$.

В результате сигнал в приемнике в соответствии с (8.3) можно записать в виде

$$s_{\text{пр}}(t) = k_0 s(t - \tau_0) + \sum_{i=1}^2 k_i s(t - \tau_a), \quad (8.6)$$

где коэффициенты k_i при $i = 0, 1, 2$ в общем случае имеют разные значения, поскольку коэффициенты отражения радиоволн на разных путях могут быть существенно разными. Второе слагаемое в (8.6) приемником с ненаправленной антенной воспринимается как одна радиоволна. Поэтому вместо (8.6) приходится записать следующее выражение:

$$s_{\text{пр}}(t) = k_0 s(t - \tau_0) + b_1 s(t - \tau_a), \quad (8.7)$$

где $b_1 = k_1 + k_2$ — коэффициент, определяющий мощность суммарной радиоволны. В соответствии с (8.6) или (8.7) принято говорить, что приемник принимает два луча, временная задержка между которыми равна $\tau = \tau_a - \tau_0$. Абсолютное значение времени задержки первого луча τ_0 в теории передачи информации не играет существенной роли, в то время как относительная задержка между лучами является важным параметром, существенным образом влияющим на качество передачи.

Поскольку в данном разделе рассматриваются радиосигналы, то вместо вещественных функций в (8.6) или в (8.7) удобнее использовать их комплексные огибающие, что позволит рассматривать более реальные условия распространения радиоволн с учетом дифракции на препятствиях, отражений с комплексным коэффициентом отражения, проникновениями через препятствия и т.д.

В общем случае при любом способе модуляции радиосигнал, излучаемый передатчиком, на любом интервале времени можно записать в виде следующей функции с вещественными значениями:

$$\begin{aligned} s(t) &= a(t) \cos[2\pi f_0 t + \varphi(t)] = \operatorname{Re} \{ a(t) \exp \{ j[2\pi f_0 t + \varphi(t)] \} \} = \\ &= \operatorname{Re} \{ a(t) \exp \{ j\varphi(t) \} \exp \{ j2\pi f_0 t \} \} = \operatorname{Re} \{ \dot{a}(t) \exp \{ j2\pi f_0 t \} \}, \end{aligned}$$

где

$$\dot{a}(t) = a(t) \exp \{ j\varphi(t) \} \tag{8.8}$$

называют *комплексной огибающей вещественного сигнала* $s(t)$. Если теперь рассмотреть данный сигнал на выходе широкополосного канала с постоянными параметрами*, т.е. на входе приемника, то в достаточно реальных условиях распространения радиоволн для него справедливо представление:

$$\begin{aligned} s(t) &= k \cdot a(t) \cos[2\pi f_0 t + \varphi(t) + \psi] = \operatorname{Re} \{ k \cdot a(t) \exp \{ j[2\pi f_0 t + \varphi(t) + \psi] \} \} = \\ &= \operatorname{Re} \{ k \exp \{ j\psi \} \cdot a(t) \exp \{ j\varphi(t) \} \cdot \exp \{ j2\pi f_0 t \} \} = \operatorname{Re} \{ \dot{k} \cdot \dot{a}(t) \exp \{ j2\pi f_0 t \} \} = \\ &= \operatorname{Re} \{ \dot{b}(t) \exp \{ j2\pi f_0 t \} \}. \end{aligned} \tag{8.9}$$

Здесь функция

$$\dot{b}(t) = \dot{k} \cdot \dot{a}(t) \tag{8.10}$$

является комплексной огибающей сигнала на выходе канала, а

$$\dot{k} = k \exp \{ j\psi \} \tag{8.11}$$

иногда называют *комплексным коэффициентом передачи радиоканала*.

Основным здесь принимается равенство (8.10), которое легко позволяет найти комплексную огибающую сигнала на выходе канала, если известен коэффициент передачи этого канала.

Теперь можно вернуться к равенству (8.6), в котором коэффициенты k_i , $i = 0, 1, 2$, допустимо считать комплексными величинами, т.е. полагают, что при распространении радиоволны вдоль каждого пути происходит изменение не только уровня сигнала, но и его фазы; форма сигнала остается неизменной; эти изменения на разных путях распространения различны. В этом случае вместо (8.6) можно записать

* Канал передачи считают широкополосным, если ширина его полосы пропускания больше ширины спектра сигнала. В этом случае канал не вносит линейных искажений формы сигнала. Предположение о постоянстве значений параметров каналов в дальнейшем будет снято.

$$s_{\text{np}}(t) = \text{Re} \left\{ \left[k_0 \exp\{j\psi_0\} \cdot a(t - \tau_0) \exp\{j\varphi(t - \tau_0)\} + \sum_{i=1}^2 k_i \exp\{j\psi_i\} \cdot a(t - \tau_a) \exp\{j\varphi(t - \tau_a)\} \right] \exp\{j2\pi f_0 t\} \right\} = \text{Re} \{ \dot{b}(t) \exp\{j2\pi f_0 t\} \}.$$

Здесь

$$\begin{aligned} \dot{b}(t) &= k_0 \exp\{j\psi_0\} \cdot a(t - \tau_0) \exp\{j\varphi(t - \tau_0)\} + \sum_{i=1}^2 k_i \exp\{j\psi_i\} \cdot a(t - \tau_a) \exp\{j\varphi(t - \tau_a)\} = \\ &= \dot{k}_0 \cdot \dot{a}(t - \tau_0) + \left[\sum_{i=1}^2 \dot{k}_i \right] \dot{a}(t - \tau_a) — \end{aligned} \quad (8.12)$$

комплексная огибающая сигнала на выходе канала передачи, который представляет собой сумму двух лучей с относительной задержкой между ними $\tau = \tau_a - \tau_0$. Мощность сигнала первого луча пропорциональна величине $|\dot{k}_0|^2$, второго — $\left| \sum_{i=1}^2 \dot{k}_i \right|^2$. Обратим внимание на то, что второй луч является суммой двух радиоволн, распространявшихся разными путями, коэффициенты передачи каналов вдоль которых различны. Величины \dot{k}_0 и $\sum_{i=1}^2 \dot{k}_i$ можно назвать комплексными амплитудами первого и второго луча сигнала в приемнике. На комплексной плоскости комплексные амплитуды можно представлять векторами, длины и углы которых определяются как модули и аргументы соответствующих величин.

Представление (8.12) для сигнала в приемнике — все же частное. В более общем случае этот сигнал — сумма большего числа лучей, каждый из которых в свою очередь является суммой многих радиоволн, имеющих одинаковое время распространения от передатчика к приемнику, хотя и прошедших разными путями. Комплексная огибающая принимаемого сигнала может быть представлена следующим выражением:

$$\dot{b}(t) = \sum_{i=0}^{n-1} \dot{k}_i \dot{a}(t - \tau_i), \quad (8.13)$$

в котором параметр n определяет число лучей. Комплексные коэффициенты \dot{k}_i каждого луча могут представлять собой сумму конечного или даже бесконечного числа комплексных величин, каждая из которых может интерпретироваться как комплексный коэффициент передачи канала вдоль соответствующего пути распространения радиоволны.

Представление комплексной огибающей сигнала на выходе радиоканала в виде (8.13) может послужить основой для описания такого канала с помощью комплексной импульсной характеристики

$$\dot{h}(\tau) = \sum_{i=0}^{n-1} \dot{k}_i \cdot \delta(\tau - \tau_i). \quad (8.14)$$

Канал с такой импульсной характеристикой обычно называют *многолучевым*. Если значения параметров \dot{k}_i , τ_i , $i = 0, 1, 2, \dots, n - 1$ и число лучей n такого канала не изменяются во времени, то такой канал называют *многолучевым с постоянными параметрами*.

Модуль импульсной характеристики (8.14) радиоканала полезно представить графически; пример соответствующего графика приведен на рис. 8.2, на котором по оси ординат откладываются значения коэффициентов $|\dot{k}_i|^2$, $i = 0, 1, 2, \dots, n - 1$, определяющих мощности

соответствующих лучей*, а по оси абсцисс — время распространения радиоволн i -го луча от передатчика до приемника.

Такое графическое представление импульсной характеристики иллюстрирует рассеяние мощности принимаемого сигнала по времени в многолучевом канале. Эффект рассеяния обуславливает существенные ограничения на достижимое качество передачи информации по многолучевому каналу и приводит к необходимости заметного усложнения алгоритмов демодуляции принимаемого сигнала в приемнике. Опыт построения систем передачи информации по многолучевым каналам позволил установить наличие зависимости многих важных показателей качества систем передачи от некоторых обобщенных параметров импульсной характеристики каналов. Одним из наиболее важных таких параметров является расширение задержки $\Delta\tau$, который можно определить как интервал времени между первым и последним лучом импульсной характеристики канала, т.е. $\Delta\tau = \tau_{n-1} - \tau_0$. Однако в действительности такое определение оказывается мало полезным по следующим причинам.

Так как с ростом времени задержки луча увеличивается длина пути, по которому распространяется соответствующая радиоволна, то мощность сигналов лучей в среднем падает с увеличением номера луча (штриховая линия на рис. 8.2); однако эта зависимость статистическая, поскольку потери распространения вдоль конкретного пути существенно зависят от наличия препятствий на данном пути. Не редки случаи, когда мощность первого принимаемого луча, прошедшего по кратчайшему пути, не является максимальной (имеет место при отсутствии прямой видимости между антеннами передатчика и приемника и наличии сильных отраженных лучей).

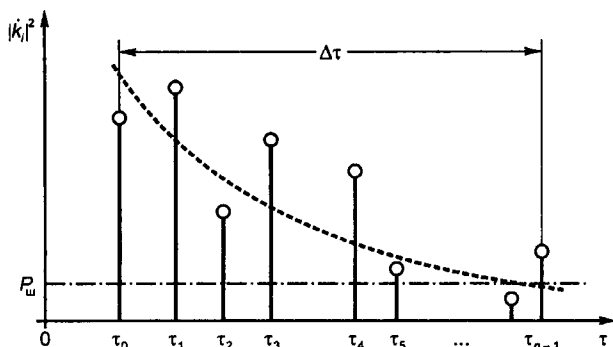


Рис. 8.2. Лучи многолучевого радиоканала

Импульсная характеристика канала передачи существенно зависит от взаимного положения передатчика и приемника на местности и может существенно изменяться при их перемещении. Не всегда можно достаточно легко указать число принимаемых лучей и выделить луч с максимальной задержкой, поскольку число лучей может быть очень большим, а мощность многих из них может оказаться сравнимой или даже меньше мощности собственного шума приемника (пунктирная линия на рис. 8.2).

Кроме того, на рис. 8.2 приведен пример так называемой *дискретной многолучевости*, когда соседние лучи имеют конечную разность хода $\tau_i - \tau_{i-1}$. В реальных радиоканалах нередки случаи, когда импульсная характеристика канала не является дискретной. На рис. 8.3 приведен график такой импульсной характеристики, который присущ интенсивной городской застройке. При изменении взаимного положения передатчика и приемника импульсная характеристика канала существенно изменяется, так что ее приходится рассматривать как случайный процесс. Следовательно, ее свойства можно характеризовать только статистическими числовыми параметрами.

* На рис. 8.2 началом отсчета времени по оси абсцисс принят момент излучения передатчиком очень короткого импульса; импульсная характеристика радиоканала является откликом канала на такое входное воздействие.

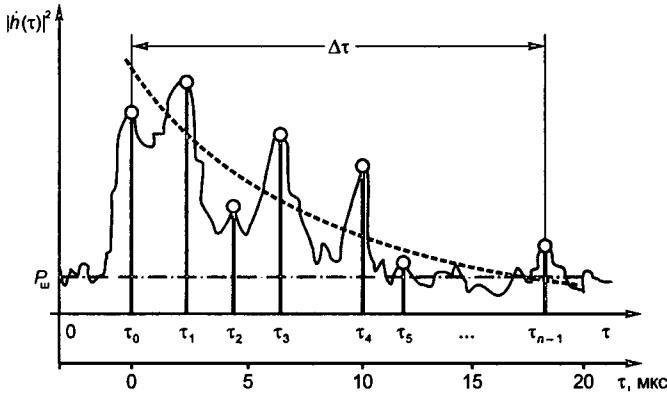


Рис. 8.3. Импульсная характеристика радиоканала в городской застройке

При разработке современных систем передачи информации по многолучевым каналам ограничиваются дискретными моделями многолучевости. При любом числе обрабатываемых в приемнике лучей выбирают дискретные лучи с наибольшей мощностью принимаемого сигнала (см. рис. 8.3). Испытания таких систем также проводят при использовании дискретных моделей радиоканалов. Например, системы стандарта GSM-900 должны испытываться при дискретных моделях

многолучевости радиоканалов, некоторые сведения о которых можно найти в 8.5.

Представление широкополосного многолучевого канала с помощью импульсного отклика эквивалентно моделированию канала соответствующим линейным фильтром. Импульсный отклик содержит полное описание фильтра, достаточное для решения любых задач, в которых такой фильтр фигурирует. В частности, могут быть решены задачи анализа качества системы передачи по такому каналу при фиксированной структуре приемника или синтеза оптимальных алгоритмов выделения полезной информации из процессов, наблюдаемых на выходе таких каналов.

Представление (8.14) для импульсной характеристики радиоканала справедливо только для узкого класса реальных радиоканалов, которые называются *каналами с постоянными параметрами*, или инвариантными во времени каналами. В действительности большая часть реальных радиоканалов не являются таковыми, поскольку их импульсная характеристика изменяется во времени. В частности, простейшее обобщение представления (8.14) состоит в том, что параметры \dot{k}_i , τ_i , $i = 0, 1, 2, \dots, n-1$ импульсной характеристики изменяются во времени, т.е. импульсная характеристика

$$\dot{h}(t, \tau) = \sum_{i=0}^{n-1} \dot{k}_i(t) \delta(\tau - \tau_i(t)) \quad (8.16)$$

оказывается функцией двух аргументов: для каждого фиксированного момента времени t' значения указанных параметров имеют некоторые фиксированные мгновенные значения $\dot{k}_i(t')$, $\tau_i(t')$, так что можно построить график импульсной характеристики по переменной τ , $0 \leq \tau < \infty$.

Параметры импульсной характеристики реального радиоканала изменяется во времени по многим причинам. Примером наиболее очевидной и наиболее часто имеющей место причиной является изменение взаимного положения на местности передатчика и приемника в системах связи с подвижными объектами, поскольку существенно изменяются пути распространения радиоволн. Как правило, эти изменения значительно медленнее, по сравнению с изменениями значений комплексной огибающей полезного сигнала, обусловленными модуляцией. Поэтому на интервале времени, длительность которого равна суммарной длительности десятков или даже сотен канальных символов, значения параметров импульсной

характеристики можно приближенно считать постоянными^{*}. Следует отметить, что в реальных условиях некоторые лучи могут исчезать, могут появляться новые лучи; это означает, что коэффициенты $k_i(t')$ могут принимать нулевые значения (число лучей n уменьшается) либо значение параметра n может увеличиваться. Подобные изменения импульсной характеристики радиоканала с переменными параметрами представлены на рис. 8.4. Здесь в качестве начала отсчета значений переменной τ принято значение τ_0 времени распространения прямого луча^{**}; штриховыми линиями изображены возможные мгновенные значения параметров $\tau_1, \tau_2, \dots, \tau_{n-1}$; в момент времени t_3 третий луч исчез. В общем случае изменения параметров импульсной характеристики приходится рассматривать как случайные, так что функции $\tau_i(t), i = 1, 2, \dots$ необходимо рассматривать как случайные процессы.

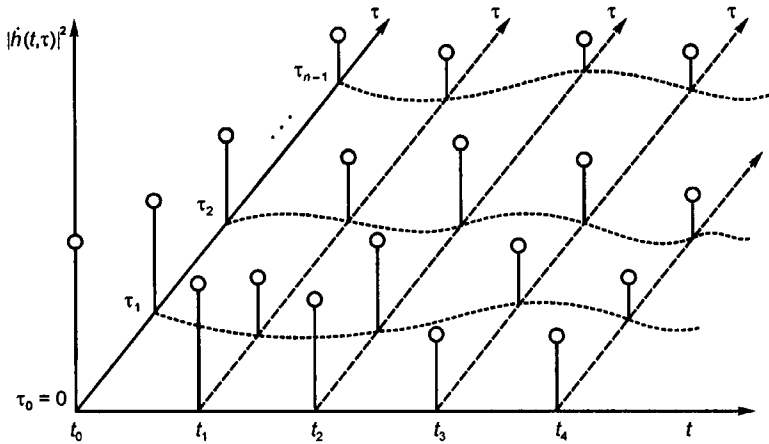


Рис. 8.4. Пример модели импульсной характеристики многолучевого радиоканала с переменными параметрами

Иногда удобно к каналу относить не только среду распространения между антеннами передатчика и приемника, но и их полосовые фильтры: формирующий фильтр основной полосы передатчика и согласованный фильтр приемника. Канал оказывается узкополосным. В общем случае импульсная характеристика такого канала с постоянными параметрами имеет следующее представление:

$$h(\tau) = h_0(\tau) \cos[2\pi f_0 \tau + \psi(\tau)], \quad 0 \leq \tau < \infty, \quad (8.17)$$

где $h_0(\tau)$ — огибающая импульсной характеристики. Для описания такой импульсной характеристики удобно ввести комплексную огибающую по аналогии с комплексной огибающей узкополосного сигнала:

$$\hat{h}(\tau) = h(\tau) \exp\{j\psi(\tau)\}, \quad 0 \leq \tau < \infty. \quad (8.18)$$

Если канал не является инвариантным по времени, то вместо (8.18) используют более сложную функцию

^{*} Справедливо при медленных замираниях.

^{**} В реальных системах, работающих с такими каналами, иногда предусматривается слежение по задержке за каждым используемым лучом. Здесь можно принять, что слежение осуществляется за первым лучом, а задержки остальных лучей измеряются по отношению к первому.

$$\dot{h}(t, \tau) = h(t, \tau) \exp\{j\psi(t, \tau)\}, \quad (8.19)$$

в которой и огибающая, и угловая модуляция отклика зависят от времени. Соответствующая дискретная модель может быть представлена выражением

$$\dot{h}(t, \tau) = \sum_{i=0}^{n-1} h_i(t, \tau) \exp\{j[2\pi f_0 \tau_i(t) + \theta_i(t, \tau)]\} \delta[\tau - \tau_i(t)], \quad (8.20)$$

где $h_i(t, \tau)$ и $\tau_i(t)$ — вещественные функции, определяющие огибающую и время задержки i -го луча в момент времени t ; фазовый множитель $\psi(t, \tau) = 2\pi f_0 \tau_i(t) + \theta_i(t, \tau)$ характеризует смещение фазы, обусловленное дополнительной задержкой i -го луча при распространении и изменением фазы в канале.

8.2. Мощность сигнала на выходе многолучевого канала

Приведем соотношения, определяющие мощность сигнала, принимаемого приемником на выходе многолучевого канала, для которого используем дискретную модель (8.20), для случая, когда передатчик излучает немодулированное несущее колебание с комплексной огибающей $\dot{a}(t) = 2$. Средняя по времени мощность такого сигнала равна единице. Комплексная огибающая сигнала на выходе канала (8.20)

$$\dot{r}(t) = \int_{-\infty}^t \dot{h}(t, \tau) \dot{a}(t - \tau) d\tau = 2 \sum_{i=0}^{n-1} h_i(t, \tau_i) \exp\{j\psi(t, \tau_i)\}. \quad (8.21)$$

Мгновенная мощность такого сигнала

$$P(t) = \frac{1}{2} |\dot{r}(t)|^2 = \frac{1}{2} \left| \sum_{i=0}^{n-1} h_i(t, \tau_i) \exp\{j\psi(t, \tau_i)\} \right|^2. \quad (8.22)$$

Амплитуды $h_i = h_i(t, \tau_i)$ лучевых компонент принимаемого сигнала как функции времени t при изменении взаимного положения передатчика и приемника обычно изменяются сравнительно медленно, в то время как фазы $\psi_i = \psi(t, \tau_i)$ этих компонент изменяются существенно даже при небольших изменениях t (или при небольших перемещениях передатчика или приемника), поскольку зависят от значения частоты несущего колебания (содержат слагаемые $2\pi f_0 \tau_i(t)$). Так как комплексные слагаемые в (8.22) складываются как вектора с учетом их фаз, то значение суммы может измениться существенно даже при незначительных изменениях расстояния между передатчиком и приемником (порядка длины волны). Таким образом, при непрерывном излучении передатчиком немодулированного несущего колебания значение огибающей сигнала на выходе канала может быстро и в широком диапазоне изменяться во времени при перемещении передатчика или приемника, приводя к глубоким замираниям. Этот эффект иллюстрирует рис. 8.5: мгновенная мощность принимаемого многолучевого сигнала может существенно изменяться при незначительном взаимном перемещении передатчика и приемника.

В представлении (8.22) векторы амплитуд $\mathbf{h} = [h_0, h_1, \dots, h_{n-1}]$ и вектор фаз $\Psi = [\psi_0, \psi_1, \dots, \psi_{n-1}]$ лучевых компонент приходится рассматривать как случайные векторы. Поэтому среднее значение мгновенной мощности для любого фиксированного момента времени можно вычислить как математическое ожидание функции $P(t; \mathbf{h}, \Psi)$ по совместному распределению случайных векторов \mathbf{h} и Ψ :

$$\begin{aligned}
 P_{cp} &= M_{h,\psi} [P(t; \mathbf{h}, \psi)] = \frac{1}{2} M_{h,\psi} \left[\sum_{i=0}^{n-1} h_i \exp\{j\psi_i\} \sum_{j=0}^{n-1} h_j \exp\{-j\psi_j\} \right] = \\
 &= \frac{1}{2} \sum_{i=0}^{n-1} M[h_i^2] + \sum_{i=0}^{n-1} \sum_{j=0, j \neq i}^{n-1} M[h_i h_j] M[\exp\{j(\psi_i - \psi_j)\}] = \frac{1}{2} \sum_{i=0}^{n-1} M[h_i^2].
 \end{aligned}
 \tag{8.23}$$

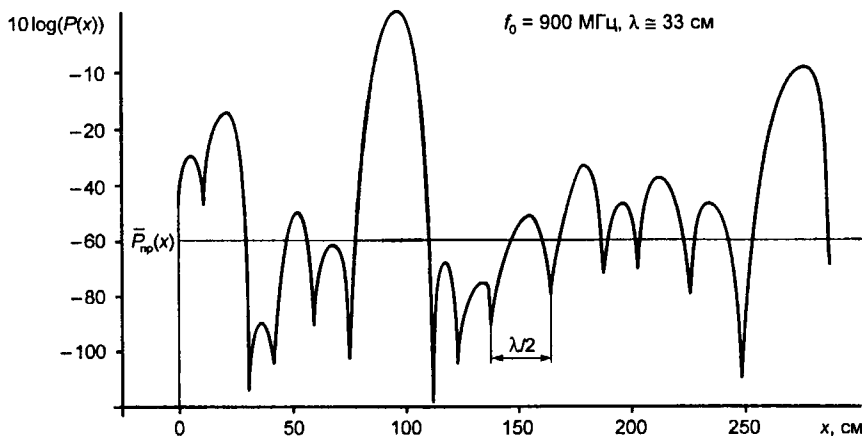


Рис. 8.5. Изменение мгновенной мощности принимаемого многолучевого сигнала при перемещении приемника

В (8.23) принято, что эти векторы взаимно независимы, что выполняется в реальных условиях. Кроме того, фазы ψ_i и ψ_j разных лучевых компонент можно считать независимыми равномерно распределенными случайными величинами, поскольку разные лучи формируются большим числом радиоволн, пришедших к приемнику по разным путям, длина которых может превышать несколько десятков или даже сотен длин волн. Поэтому в (8.23) все слагаемые $M[\exp\{j(\psi_i - \psi_j)\}] = 0$ при $i \neq j$, так что среднее значение мощности многолучевого сигнала равно сумме средних мощностей $\frac{1}{2} M[h_i^2]$ сигналов всех принимаемых лучей*, $i = 0, 1, 2, \dots, n - 1$. Отметим также, что при вычислениях в (8.23) принято, что распределение вектора \mathbf{h} не зависит от времени, вследствие чего среднее значение мгновенной мощности принимаемого многолучевого канала не зависит от времени.

Рассмотрим теперь случай, когда многолучевой канал зондируется широкополосным радиосигналом

$$s(t) = \text{Re}\{2\dot{a}(t) \exp\{j2\pi f_0 t\}\}, \quad 0 \leq t \leq T_c, \tag{8.24}$$

длительность которого T_c намного меньше любого временного интервала $\Delta\tau_i = \tau_i - \tau_{i-1}$, $i = 0, 1, 2, \dots, n - 1$ между лучами, а огибающая $|\dot{a}(t)| = 1$ на этом интервале. Средняя мощность такого сигнала на интервале $0 \leq t \leq T_c$ равна 1. Комплексная огибающая сигнала на

* Не прибегая к аккуратному математическому анализу, этот результат можно обосновать следующим образом: если лучи независимы, то их мощности складываются, так как математическое ожидание суммы случайных величин всегда равно сумме их математических ожиданий.

выходе канала с импульсной характеристикой (8.20) определяется формулой (8.21) и в данном примере будет иметь вид

$$\dot{r}(t) = \frac{1}{2} \int_{-\infty}^t \dot{h}(t, \tau) 2\dot{a}(t - \tau) d\tau = \sum_{i=0}^{n-1} h_i(t, \tau_i) \dot{a}(t - \tau_i) \exp\{j\psi(t, \tau_i)\}. \quad (8.25)$$

Теперь слагаемые в формуле (8.25) отличны от нуля на непересекающихся интервалах времени. Исследуем произвольный момент времени t_0 и найдем среднюю мощность этого сигнала на локальном интервале времени, содержащем t_0 , длительность которого равна расширению задержки $\Delta\tau$ рассматриваемого канала и который содержит все лучевые компоненты. На данном интервале амплитуды лучей $h_i(t_0) = h_i(t_0, \tau_i)$ можно считать постоянными, а фазовые множители $\psi_i = \psi(t, \tau_i)$ — постоянными на коротких интервалах длительностью T_c . Тогда средняя мощность сигнала (8.25) на локальном интервале времени для произвольного момента времени t_0

$$\begin{aligned} \bar{P}(t_0; \mathbf{h}) &= \frac{1}{\Delta\tau} \int_{t_0}^{t_0 + \Delta\tau} \frac{1}{2} |\dot{r}(t)|^2 dt = \frac{1}{2\Delta\tau} \int_{t_0}^{t_0 + \Delta\tau} \left| \sum_{i=0}^{n-1} h_i(t_0) \dot{a}(t - \tau_i) \exp\{j\psi_i\} \right|^2 dt = \\ &= \frac{1}{2\Delta\tau} \sum_{i=0}^{n-1} \int_{t_0}^{t_0 + T_c} h_i^2(t_0) dt = \frac{nT_c}{2\Delta\tau} \sum_{i=0}^{n-1} h_i^2(t_0) \end{aligned} \quad (8.26)$$

равна суммарной мощности лучевых компонент, нормированной на отношение суммарной длительности лучевых компонент nT_c к расширению задержки $\Delta\tau$ канала. Черта над буквой P здесь обозначает среднее значение по времени.

В рассматриваемом случае широкополосного сигнала на входе канала лучевые компоненты на выходе канала полностью разрешены (не перекрываются во времени). Так как амплитуда каждой лучевой компоненты меняется во времени сравнительно медленно, то и суммарная мгновенная мощность всех лучевых компонент на любом локальном интервале времени сравнительно медленно изменяется во времени.

Поскольку амплитуды лучей h_i , $i = 0, 1, \dots, n-1$ являются случайными величинами, то можно найти математическое ожидание среднего по времени значения мгновенной мощности

$$\bar{P}_{cp}(t_0) = M_{\mathbf{h}} \{ \bar{P}(t_0; \mathbf{h}) \} = M_{\mathbf{h}} \left\{ \frac{nT_c}{2\Delta\tau} \sum_{i=0}^{n-1} h_i^2(t_0) \right\} = \frac{nT_c}{2\Delta\tau} \sum_{i=0}^{n-1} M_{h_i} \{ h_i^2(t_0) \}. \quad (8.27)$$

Если распределения амплитуд лучей не зависят от времени, то из (8.27) получаем

$$\bar{P}_{cp} = \frac{nT_c}{\Delta\tau} \frac{1}{2} \sum_{i=0}^{n-1} M_{h_i} \{ h_i^2 \}, \quad (8.28)$$

что с точностью до постоянного множителя совпадает с (8.23).

Таким образом, средняя мощность сигнала на выходе многолучевого канала при излучении передатчиком как узкополосного, так и широкополосного сигналов равна сумме средних мощностей сигналов всех принимаемых лучей (при статистической независимости лучей). Однако мгновенная мощность многолучевого сигнала при узкополосном сигнале на входе канала испытывает быстрые и глубокие замирания, в то время как при широкополосном сигнале на входе эти изменения медленные и незначительные.

В современных системах передачи информации по радиоканалам предусматривается возможность измерения или прогноза импульсной характеристики. Для этих целей канал может зондироваться коротким импульсным сигналом, длительность которого должна быть значительно меньше значения расширения задержки $\Delta\tau$. Поскольку для любого момента времени t импульсная характеристика радиоканала является непрерывной функцией по

переменной τ , то для построения дискретной модели импульсной характеристики ось значений τ дискретизируется на интервале $\Delta\tau$ с постоянным шагом $\delta\tau$. Число учитываемых лучей может быть оценено значением $\Delta\tau/(\delta\tau) = n - 1$. Далее можно оценивать значения импульсной характеристики для значений $\tau_i = i \cdot \delta\tau, i = 0, 1, 2, \dots, n - 1$. Существуют и так называемые *слепые методы* оценивания импульсной характеристики радиоканала, когда в качестве зондирующего сигнала используется информационный сигнал.

Если сигнал, используемый для зондирования радиоканала, имеет длительность, значительно меньше расширения задержки $\Delta\tau$, то можно определить многолучевой профиль задержек

$$P(t, \tau) \approx k |h(t, \tau)|^2, \tag{8.29}$$

характеризующий распределение мощности принимаемого сигнала по переменной τ для каждого момента времени t , т.е. распределение мощности между лучевыми компонентами на выходе канала. Для различных условий передачи информации (для различных каналов, различных взаимных положений передатчика и приемника и т.д.) можно получить локальные средние по времени значения профилей задержек $P(\tau)$, наиболее характерные для данных условий и инвариантные по времени. Коэффициент k в (8.29) можно представить как отношение мощности зондирующего сигнала на входе канала к суммарной мощности принимаемого многолучевого сигнала.

Один из возможных способов экспериментального определения импульсной характеристики многолучевого канала состоит в зондировании канала сигналом с расширенным спектром. Расширение спектра осуществляется периодической псевдослучайной последовательностью, а прием многолучевого сигнала обеспечивается приемником с согласованным фильтром. Функциональные схемы передатчика и приемника для такого эксперимента представлены на рис. 8.6. Частота следования элементарных символов последовательности $F_{cn} = 1/\tau_0$; спектральную плотность мощности излучаемого сигнала можно приближенно выразить соотношением

$$S(f) = k \left[\frac{\sin(\pi(f - f_0)\tau_0)}{\pi(f - f_0)\tau_0} \right]^2 \tag{8.30}$$

при ширине спектра между нулевыми значениями $2\Delta F = 2/\tau_0$.

После прохождения через радиоканал сигнал подвергается полосовой фильтрации в приемнике в фильтре с полосой пропускания $2\Delta F$, а затем сворачивается по спектру в корреляторе с помощью генератора псевдослучайной последовательности, аналогичной той, которая используется в передатчике, но с несколько меньшей частотой тактового генератора $F'_{cn} = 1/\tau'_0$. При таких частотах тактовых генераторов передатчика и приемника более быстрая последовательность передатчика со скоростью $1/(F_{cn} - F'_{cn})$ перемещается относительно последовательности приемника. Если эти последовательности смещены друг относительно друга, то спектр сигнала передатчика будет дополнительно расширяться. В тот момент, когда коды последовательностей одного луча принимаемого сигнала и приемника практически совпадут, на выходе перемножителя приемника будет сформирован узкополосный радиоимпульс, который будет выделен полосовым фильтром приемника с полосой пропускания $2\Delta F \approx 2(F_{cn} - F'_{cn})$. Поскольку представленный на рис. 8.6 приемник является линейным, то подобным преобразованиям будут подвергнуты независимо друг от друга все имеющиеся на выходе радиоканала сигналы всех лучей.

* Такие операции, например, предусматриваются рекомендациями стандарта GSM-900.

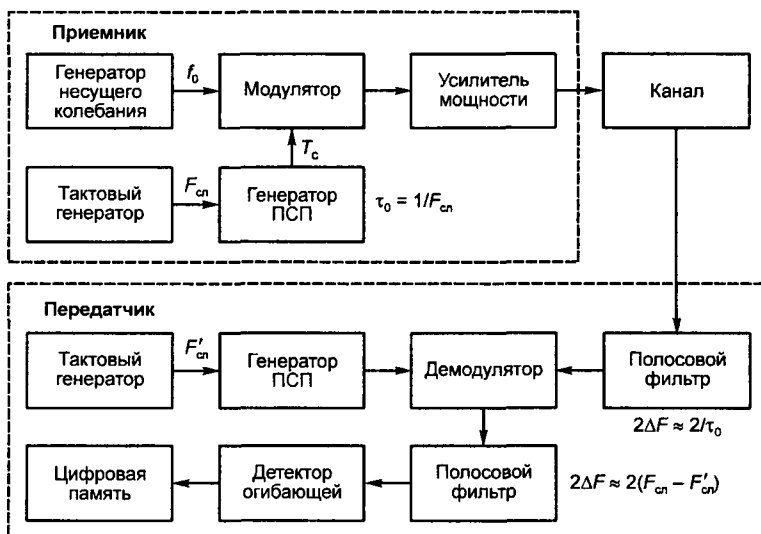


Рис. 8.6. Система измерения импульсной характеристики многолучевого канала (разрешающая способность по времени равна длительности элементарного импульса τ_0)

При таком способе зондирования канала интервал времени между двумя соседними максимальными значениями корреляции периодических последовательностей передатчика и приемника в однолучевом канале

$$\Delta T = \tau_0 L \frac{F_{cn}}{F_{cn} - F'_{cn}}, \quad (8.31)$$

где L — число элементарных символов псевдослучайной последовательности. Отношение $\frac{F_{cn}}{F_{cn} - F'_{cn}} = \gamma$, называемое *коэффициентом скольжения*, определяет эквивалентный масштаб времени измерения:

$$\text{Время измерения} = (\text{Реальный интервал времени}) \cdot \gamma. \quad (8.32)$$

Длина последовательности должна быть выбрана таким образом, чтобы обеспечить однозначное измерение времени прихода луча с максимальным временем задержки τ_{\max} относительно прямого луча. Для данного периода последовательности должен удовлетворять условию $\tau_0 L > \tau_{\max}$. Период последовательности определяет также зону на местности, из которой могут быть приняты лучевые компоненты; для этого достаточно период последовательности умножить на скорость распространения электромагнитной волны.

Способ зондирования многолучевого канала сигналом с расширенным спектром обладает определенными преимуществами по сравнению с зондированием коротким одиночным импульсом: обеспечивает подавление узкополосных помех, расширение зоны исследования на местности при фиксированной мощности передатчика; существенно снижаются требования к динамическому диапазону передатчика и приемника; использование этого метода исключает необходимость синхронизации передатчика и приемника; чувствительность метода легко регулируется путем изменения коэффициента скольжения γ и полосы пропускания узкополосного фильтра коррелятора. Эти преимущества сопровождаются и

некоторыми недостатками: результаты измерений получаются не в реальном времени; время измерений значительно возрастает; используется некогерентный прием, так что фазы отдельных лучей не измеряются.

Прямое преобразование Фурье импульсной характеристики многолучевого канала с постоянными параметрами

$$\dot{K}(f) = \int_0^{\infty} h(\tau) e^{-j2\pi f\tau} d\tau, \quad -\infty < f < +\infty \quad (8.33)$$

определяет комплексный коэффициент передачи радиоканала, для которого также предложено несколько способов экспериментального измерения. Обратное преобразование Фурье

$$h(\tau) = \int_{-\infty}^{+\infty} \dot{K}(f) e^{j2\pi f\tau} df, \quad 0 \leq \tau < +\infty \quad (8.34)$$

позволяет найти импульсную характеристику.

Измерения характеристики (8.33) осуществляется на некоторой сетке частот, расположенной симметрично относительно частоты несущего колебания системы передачи. Число частот и расстояния между ними определяют разрешающую способность во временной области при нахождении импульсной характеристики канала. На каждой частоте на вход канала подается гармонический сигнал с известным значением амплитуды $A_{\text{вх}}$; измеряется амплитуда $A_{\text{вых}}$ сигнала на выходе канала и его фаза $\Phi_{\text{вых}}$ относительно фазы сигнала на входе; коэффициент передачи

$$\dot{K}(f_i) = \frac{A_{\text{вых}}(f_i)}{A_{\text{вх}}(f_i)} e^{j\Phi_{\text{вых}}(f_i)}, \quad i = 0, 1, 2, \dots, n - 1. \quad (8.35)$$

Обратное дискретное преобразование Фурье обеспечивает возможность вычисления импульсной характеристики канала.

Такой способ измерения импульсной характеристики удобен тем, что позволяет получить оценку не только модуля коэффициента передачи ограниченного по полосе канала, но и соответствующую ему фазу. Однако требуется достаточно точная калибровка измерительного оборудования и жесткая синхронизация передатчика и приемника, что можно осуществить только при небольших расстояниях между передатчиком и приемником. Еще одним недостатком этого способа измерения является то, что измерения выполняются не в реальном времени. Для каналов с переменными во времени параметрами комплексный коэффициент передачи может изменяться во времени, приводя к значительным ошибкам измерения импульсной характеристики.

8.3. Числовые характеристики многолучевых каналов

8.3.1. Рассеяние во времени

Импульсная характеристика канала характеризует рассеяние мощности принимаемого сигнала по времени. Основная причина этого эффекта — естественна и вполне очевидна — наличие в радиоканале путей распространения радиоволн с различной длиной с отражениями и рассеянием. При проектировании и анализе различных радиосистем, использующих многолучевые каналы, полезно иметь числовые параметры, которые давали бы укрупненное описание каналов.

Наиболее важными параметрами импульсной характеристики радиоканала, характеризующими рассеяние во времени, признаются среднее время задержки и среднеквадратическое расширение задержки, которые определяются с помощью следующих соотношений:

$$\tau_{\text{cp}} = \frac{\int_0^{\infty} \tau |\dot{h}(\tau)|^2 d\tau}{\int_0^{\infty} |\dot{h}(\tau)|^2 d\tau}; \quad \sigma_{\tau} = \sqrt{M[\tau^2] - (\tau_{\text{cp}})^2}, \quad (8.36)$$

$$\text{где } M[\tau^2] = \frac{\int_0^{\infty} \tau^2 |\dot{h}(\tau)|^2 d\tau}{\int_0^{\infty} |\dot{h}(\tau)|^2 d\tau}.$$

Здесь в качестве начала отсчета времени задержки принят момент времени прихода первого луча. При практическом определении значений параметров учитывают наличие шума на входе приемника и необходимость обеспечения определенного превышения мгновенной мощности принимаемого луча некоторого порогового значения $P_{\text{пор}}$, которое должно превышать мощность присутствующего шума; в результате интервалы интегрирования в (8.36) выбираются конечными. Соответствующий пример для радиоканала внутри помещения и реальные числовые значения приведены на рис. 8.7, при построении которого использованы числовые данные из [8.1, с. 163].

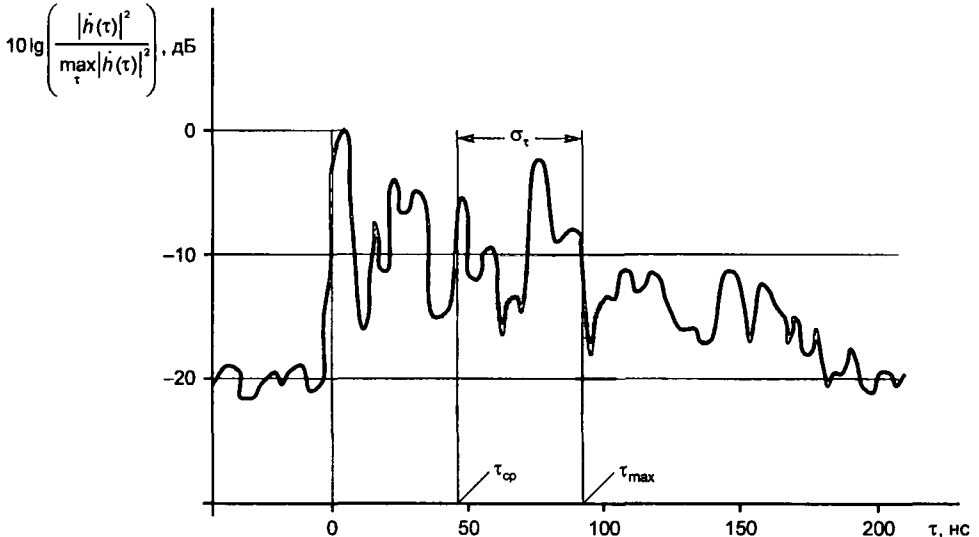


Рис. 8.7. Пример измерений рассеяния мощности сигнала во времени для радиоканала внутри здания ($\sigma_{\tau} = 46$ нс; $\tau_{\text{cp}} = 45$ нс; $\tau_{\text{max}} = 84$ нс при $P_{\text{пор}} = -10$ дБ)

В литературе опубликованы многочисленные результаты экспериментальных измерений указанных параметров радиоканалов в различных условиях. В табл. 8.1 приведены некоторые результаты статистической обработки таких измерений [8.1].

Таблица 8.1. Результаты экспериментальных измерений рассеяния задержки

Условия измерения	Частота, МГц	Рассеяние задержки σ_τ , нс	Примечание
Город	910	Среднее значение 1300 Среднеквадратическое отклонение 600 Максимальное значение 3500	г. Нью-Йорк
	892	10000–25000	г. Сан-Франциско (наихудшие случаи)
Пригород	910	200–310	Средний типичный случай
	910	1960–2110	Средний наихудший случай
Внутри здания	850	Максимальное значение 270	Офис компании
	1900	Среднее значение 70–94 Максимальное значение 1470	г. Сан-Франциско

8.3.2. Полоса когерентности по частоте

Полоса когерентности по частоте как параметр радиоканала вводится на основе статистической связи с другим параметром — среднеквадратическим расширением задержки, введенным в предыдущем разделе. Наличие связи между значениями этих параметров подтверждается экспериментально, однако установить аналитическую связь между ними не удалось. Поэтому, основываясь на экспериментальных данных, для практического использования рекомендуются некоторые приближенные соотношения.

Полоса когерентности представляет собой статистическую количественную меру интервала частот радиоканала, в которой канал может рассматриваться как «гладкий». В данной полосе канал имеет постоянный, не зависящий от частоты, модуль коэффициента передачи и линейную фазовую характеристику; все спектральные компоненты сигнала, попадающие в данную полосу, передаются по каналу с одинаковым коэффициентом усиления или ослабления и имеют одинаковое время задержки. Термин «полоса когерентности», сохраняет смысл и для каналов с переменными параметрами. Поэтому смысл термина можно пояснить и иными словами: полоса когерентности это интервал частот, в котором любые две спектральные компоненты прошедшего через канал сигнала имеют высокий коэффициент корреляции амплитуд. Два гармонических сигнала, разница частот которых больше полосы когерентности канала, подвергаются существенно разным преобразованиям в канале.

Если полосу когерентности определить как интервал частот, в котором любые две спектральные компоненты сигнала имеют коэффициент корреляции выше 0,9, то ее приближенно можно записать в виде соотношения [8.2]:

$$F_x = 1/(50\sigma_\tau). \tag{8.37}$$

Если в качестве минимально допустимого значения коэффициента корреляции принять 0,5, то полоса когерентности будет характеризоваться другим приближенным равенством:

$$F_x = 1/(5\sigma_\tau). \tag{8.38}$$

Наличие взаимосвязи между полосой когерентности и среднеквадратическим расширением задержки можно пояснить следующим образом. Изменение разности фаз между спектральными компонентами лучей с частотой f_1 , вызванное разностью $\Delta\tau$ задержек лучей в канале, равно $\Delta\varphi_1 = 2\pi f_1 \Delta\tau$ радиан, а для спектральных компонент с частотой f_2 — $\Delta\varphi_2 = 2\pi f_2 \Delta\tau$. Поэтому изменение разности фаз этих спектральных компонент $|\Delta\varphi| = 2\pi|f_1 - f_2| \Delta\tau = 2\pi \Delta f \Delta\tau$.

Сдвиг существенно влияет на результат сложения этих спектральных компонент лучей, когда, например, он достигает значения $\pi/2$ (критерий Релея). Отсюда полоса когерентности $\Delta f = 1/4\Delta t$, что близко к соотношению (8.38).

Важно отметить, что между параметрами F_k и σ_t не существует функциональной связи, а соотношения (8.37) и (8.38) являются усредненными. В общем случае для получения более точного соотношения для конкретного радиоканала необходимо проводить специальное исследование с помощью методов спектрального анализа и статистического моделирования для изучения эффектов воздействия изменяющейся во времени многолучевости канала на конкретный используемый для передачи сигнал. Поэтому при проектировании модемов для конкретных радиоканалов необходимо внимательно подходить к выбору адекватных моделей многолучевых каналов.

8.3.3. Рассеяние по частоте

Расширение задержки и полоса когерентности — параметры, посредством которых описываются дисперсионные свойства канала во времени в условиях, когда передатчик, приемник и окружающие их объекты среды распространения радиоволн неподвижны. Эти свойства сохраняются и в динамическом режиме, когда передатчик, приемник или отражатели и рассеиватели радиоволн передвигаются в пространстве. Правда, в таком режиме дисперсионные свойства реальных радиоканалов могут существенно изменяться во времени. Однако в динамическом режиме возникают и новые эффекты, о которых данные параметры не содержат никакой информации. В первую очередь здесь следует назвать эффект доплеровского расширения, обусловленный изменениями во времени свойств радиоканала в динамическом режиме.

Мгновенная частота принимаемого колебания $f_{\text{пр}} = f_0 + \frac{v \cos \theta}{\lambda}$ может изменяться при перемещении передатчика, одновременном перемещении передатчика и приемника, перемещении отражателей, окружающих передатчик и приемник. Во всех случаях изменение частоты пропорционально радиальной скорости $v_{\text{рад}} = v \cos \theta$ сближения или удаления приемника и источника излучения. Здесь v — абсолютная скорость перемещения; θ — угол между направлением перемещения и направлением на передатчик; λ — длина волны при частоте несущего колебания f_0 .

В реальных условиях распространение радиоволн происходит по многим путям, в результате чего на приемную антенну падает много плоских волн с различными значениями угла θ ; поэтому сигналы некоторых волн будут иметь положительное значение доплеровского смещения, в то время как могут присутствовать и волны с отрицательными значениями смещения частоты. В результате принимаемый сигнал, представляющий собой сумму разных волн, будет иметь спектральную плотность мощности, отличную от нуля в диапазоне частот от $f_0 - F_{\text{д max}}$ до $f_0 + F_{\text{д max}}$ даже в том случае, когда передатчик излучает немодулированное несущее колебание, т.е. когда спектральная плотность мощности излучаемого сигнала представляется лишь одной спектральной линией. Величина $\Delta F_{\text{д}} = 2F_{\text{д max}}$ называется *доплеровским расширением спектра* и является важным параметром многолучевого канала.

Таким образом, доплеровское расширение $\Delta F_{\text{д}}$ признается количественной мерой расширения спектра сигнала на выходе канала, обусловленного изменениями во времени свойств радиоканала, и определяется как полоса частот, в которой спектральная плотность мощности доплеровского спектра существенно отлична от нуля. Простые вычисления

позволяют указать, что для частоты несущего колебания около 2 ГГц и скорости перемещения объектов до 200 км/ч доплеровское расширение не превышает 500 Гц. Поэтому для большинства радиосистем доплеровское расширение пренебрежимо мало по сравнению шириной спектра полезного сигнала и на структуру линейной части приемника практически не оказывает влияния. Однако от значения данного параметра существенно зависят свойства радиоканала, которые, в свою очередь, приводят к существенному усложнению алгоритмов обработки сигнала при его демодуляции. Поэтому, как и в предыдущем разделе, при проектировании реальных радиосистем следует очень осторожно подходить к выбору численного значения данного параметра для конкретных радиоканалов.

8.3.4. Время когерентности

Время когерентности $T_{\text{ког}}$ сигнала на выходе радиоканала является параметром, который характеризует скорость изменения свойств канала во временной области, представляет собой временную характеристику канала, и в этом смысле она дуальна доплеровскому расширению, которое характеризует изменчивость канала во времени в частотной области. Можно принять, что параметры связаны обратно пропорциональной зависимостью

$$T_{\text{ког}} = 2/\Delta F_{\text{д}}. \tag{8.39}$$

Время когерентности признается статистической количественной мерой длины интервала времени, на котором импульсная характеристика радиоканала остается практически не изменной, т.е. для любого момента времени из этого интервала импульсная характеристика одна и та же. Иными словами, время когерентности — это интервал времени, на котором значения огибающей сигнала имеют высокий коэффициент корреляции. Например, время когерентности можно определить как интервал времени, на котором коэффициент корреляции двух мгновенных значений огибающей сигнала превышает значение 0,5. Результаты измерений приводят к следующему приближенному соотношению [8.1]:

$$T_{\text{ког}} = \frac{18}{16\pi\Delta F_{\text{д}}}. \tag{8.40}$$

Практические измерения показывают, что на интервале (8.39) канал обуславливает достаточно существенные изменения огибающей сигнала, а оценка (8.40) приводит к заниженному значению длительности интервала когерентности. При проектировании современных цифровых систем часто используется оценка, равная геометрическому среднему указанных величин:

$$T_{\text{ког}} = \left[\frac{36}{16\pi(\Delta F_{\text{д}})^2} \right]^{1/2} \approx \frac{0,846}{\Delta F_{\text{д}}}. \tag{8.41}$$

Время когерентности определяет такой интервал времени, что для двух моментов времени, отстоящих друг от друга на больший интервал, воздействие канала на сигнал оказывается разным. Например, если приемник движется с скоростью 90 км/ч и использует частоту несущего колебания 900 МГц, заниженная оценка (8.40) приводит к значению $T_{\text{ког}} = 2,22$ мс, в то время как из (8.41) получаем $T_{\text{ког}} = 6,77$. На указанных временных интервалах взаимное движение передатчика и приемника не приводит к изменениям свойств канала и, следовательно, к искажениям сигнала; однако искажения сигнала могут иметь место из-за многолучевого расширения времени задержки, зависящего от импульсной характеристики канала.

Часто прием сигнала осуществляется не на одну антенну, как это предполагалось выше, а на несколько с последующей совместной обработкой сигналов с выходов этих антенн.

В этом случае полезной характеристикой радиоканала служит интервал корреляции по пространству, для которого можно ввести следующее определение: это расстояние между точками приема, при котором коэффициент корреляции значений огибающих принимаемых сигналов в совпадающие моменты времени не превышает заданное значение (например, 0,1).

8.4. Вероятностные модели сигналов на выходе многолучевых каналов

8.4.1. Модель Кларка гладких замираний

К настоящему времени предложено несколько статистических моделей принимаемого сигнала, использующих различные принципы объяснения замираний в радиоканалах. Названные модели необходимы как для синтеза оптимальных алгоритмов демодуляции в таких каналах, так и для количественного анализа качества создаваемых систем передачи. Важно, что многие из этих моделей реализованы в виде специальных прикладных программ для компьютеров поскольку в современных цифровых системах передачи используется столь сложная обработка принимаемого сигнала, что оценка качества системы оказывается возможной только при статистическом имитационном компьютерном моделировании. В данном разделе приводится описание одной из наиболее распространенных моделей радиоканалов, которая широко используется многими фирмами, производителями телекоммуникационного оборудования. Данная модель называется *моделью Кларка* (Clarke R.H.) [8.1].

В такой модели статистические характеристики радиоволн сигнала, принимаемого подвижным приемником, формируются в основном эффектами рассеяния. Предполагается, что передатчик является неподвижным, имеет антенну, создающую поле с вертикальной поляризацией, излучает немодулированное гармоническое колебание с частотой f_0 Гц. Поле, падающее на антенну приемника, представляет собой сумму большого числа плоских волн с различными азимутальными углами прихода, произвольными фазами, одинаковыми средними значениями амплитуд. Предполагается, что радиоволна прямой видимости отсутствует: это соответствует условиям приема в интенсивной городской застройке.

Угол азимута α прихода одной плоской радиоволны измеряется относительно направления движения приемника, который перемещается на плоскости со скоростью v . Каждая волна испытывает доплеровское смещение, обусловленное радиальной скоростью перемещения приемника вдоль пути ее прихода. Все плоские волны приходят к антенне приемника в одно и то же время, т.е. расширения задержки времени прихода нет и имеют место гладкие замирания. Для волны с номером n путь прихода имеет азимут α_n , а доплеровское смещение, Гц, составляет

$$F_n = \frac{v}{\lambda} \cos(\alpha_n), \quad (8.42)$$

где λ — длина волны.

Каждая вертикально поляризованная плоская волна, падающая на приемную антенну, имеет E и H компоненты; электрическая компонента определяется следующим равенством:

$$E_z(t) = E_0 \sum_{n=1}^N A_n \cos[2\pi(f_0 + F_{n,n})t + \varphi_n], \quad (8.43)$$

где E_0 — среднее значение амплитуды электрической компоненты суммарного поля, которая предполагается постоянной вещественной величиной; A_n — вещественные независимые

случайные величины, представляющие собой амплитуды отдельных плоских волн; φ_n — фазы плоских волн, которые предполагаются независимыми и равномерно распределенными в интервале $[0, 2\pi]$ случайными величинами. Амплитуды плоских волн предполагаются нормированными таким образом, что сумма математических ожиданий

$$\sum_{n=1}^N M[A_n^2] = 1. \tag{8.44}$$

Число слагаемых N в (8.43) предполагается достаточно большим, а азимутальные углы равномерно распределены в интервале $[0, 2\pi]$.

Так как доплеровские смещения волн невелики по сравнению с частотой несущего колебания, то функцию времени (8.43) можно рассматривать как узкополосный случайный процесс. Кроме того, при большом числе слагаемых этот процесс будет практически гауссовским. Такой процесс можно представить в виде суммы синфазной и квадратурной компонент:

$$E_z(t) = I_c(t) \cos(2\pi f_0 t) - I_s(t) \sin(2\pi f_0 t), \tag{8.45}$$

где синфазная компонента

$$I_c(t) = E_0 \sum_{n=1}^N A_n \cos(2\pi F_{d_n} t + \varphi_n), \tag{8.46}$$

и квадратурная

$$I_s(t) = E_0 \sum_{n=1}^N A_n \sin(2\pi F_{d_n} t + \varphi_n). \tag{8.47}$$

Оба процесса (8.46) и (8.47) являются узкополосными гауссовскими процессами, независимыми друг от друга в произвольные моменты времени, имеющие нулевые средние значения и одинаковые дисперсии

$$M[I_c^2(t)] = M[I_s^2(t)] = M[E_z^2(t)] = \frac{E_0}{2}. \tag{8.48}$$

Процесс (8.45) может быть представлен также в виде квазигармонического колебания

$$E_z(t) = E_{or}(t) \cos[2\pi f_0 t - \Phi(t)], \tag{8.49}$$

где для огибающей $E_{or}(t)$ и фазы $\Phi(t)$ справедливы представления:

$$E_{or}(t) = \sqrt{I_c^2(t) + I_s^2(t)}; \tag{8.50}$$

$$\Phi(t) = \text{arccoor} \left[\frac{I_s(t)}{I_c(t)} \right]. \tag{8.51}$$

Можно показать, что одномерные плотности вероятности этих узкополосных процессов имеют вид

$$w(e_{or}) = \begin{cases} \frac{e_{or}}{\sigma^2} \exp \left\{ -\frac{e_{or}^2}{2\sigma^2} \right\}, & e_{or} \geq 0; \\ 0, & e_{or} < 0, \end{cases} \tag{8.52}$$

где $\sigma^2 = E_0^2/2$ и

$$w(\varphi) = \begin{cases} 1/(2\pi), & 0 \leq \varphi \leq 2\pi; \\ 0, & \text{вне этого интервала.} \end{cases} \tag{8.53}$$

8.4.2. Доплеровский спектр в модели Кларка

В модели сигнала, предложенной Кларком, учитывающей взаимное перемещение приемника и передатчика, предполагается, что передатчик излучает немодулированное несущее колебание с частотой f_0 . Спектральная плотность мощности такого сигнала является дискретной и содержит только одну дельта-функцию на частоте f_0 . Поскольку в радиоканале имеет место эффект доплеровского расширения спектра, то сигнал, принимаемый движущимся приемником, будет иметь непрерывный спектр, который можно назвать доплеровским; его форма будет определена в данном разделе.

Рассмотрим случай, когда в (8.43) $N \rightarrow \infty$, т.е. число приходящих плоских волн неограниченно возрастает, угловое расстояние между соседними плоскими волнами и мощность каждой плоской волны убывают так, что суммарная мощность волн, приходящих в азимутальном угле $\Delta\alpha$ из любого направления α равна $p(\alpha)\Delta\alpha$. Если характеристика направленности антенны приемника в азимутальной плоскости есть $G(\alpha)$, то суммарная мощность принимаемого сигнала

$$P_{\text{пр}} = \int_0^{2\pi} G_0 G(\alpha) p(\alpha) d\alpha, \quad (8.54)$$

где G_0 — коэффициент усиления антенны приемника относительно изотропного излучателя*.

Так как приемник перемещается со скоростью v , то компонента сигнала, приходящая на приемную антенну из угла $d\alpha$, около направления α будет иметь мощность $G_0 G(\alpha) p(\alpha) d\alpha$ и частоту

$$f = f_0 + \frac{v}{\lambda} \cos(\alpha) = f_0 + \Delta F_{\text{д max}} \cos(\alpha). \quad (8.55)$$

Обратим внимание, что функция (8.55) является симметричной по переменной α . Это означает, что две компоненты сигнала, приходящие с разных направлений α и $-\alpha$, будут иметь одну ту же частоту $f(\alpha) = f(-\alpha) = f$. Следовательно, мощность сигнала в узкой полосе df , соответствующей углу $d\alpha$, около частоты f , соответствующей направлениям α и $-\alpha$, определится равенством

$$S(f(\alpha))df = G_0 [G(\alpha)p(\alpha) + G(-\alpha)p(-\alpha)]d\alpha, \quad (8.56)$$

где $0 \leq \alpha \leq 2\pi$. Из (8.56) можно найти значение спектральной плотности мощности сигнала, принимаемого движущимся приемником:

$$S(f(\alpha)) = G_0 [G(\alpha)p(\alpha) + G(-\alpha)p(-\alpha)] \left| \frac{d\alpha}{df} \right|. \quad (8.57)$$

Теперь осталось найти производную $d\alpha/df$ из (8.55). Так как

$$\alpha = \arccos \left(\frac{f - f_0}{\Delta F_{\text{д max}}} \right), \text{ то } \frac{d\alpha}{df} = - \frac{1}{\Delta F_{\text{д max}} \sqrt{1 - \left(\frac{f - f_0}{\Delta F_{\text{д max}}} \right)^2}}.$$

Подставляя полученное выражение в (8.57), окончательно имеем

$$S(f(\alpha)) = \frac{G_0 [G(\alpha)p(\alpha) + G(-\alpha)p(-\alpha)]}{\Delta F_{\text{д max}} \sqrt{1 - \left(\frac{f(\alpha) - f_0}{\Delta F_{\text{д max}}} \right)^2}}, \quad 0 \leq \alpha \leq 2\pi, \quad (8.58)$$

где зависимость $f(\alpha)$ определяется (8.55).

* Для четвертьволновой вертикальной антенны $G_0 = 1,5$. Другие сведения об антеннах приведены в гл. 1.

Формула (8.58) не очень удобна для практического использования, поскольку основным аргументом функций в ней является направление прихода радиоволны α , в то время как частота f сигнала, принимаемого с этого направления, оказывается функцией от α . Кроме того, антенна подвижного приемника в азимутальной плоскости имеет круговую характеристику направленности, поскольку прием должен быть обеспечен при любом взаимном положении приемника и передатчика. Следовательно, в (8.58) можно принять $G(\alpha) = 1$. Дополнительное упрощение (8.58) возможно путем введения предположения, что мощность принимаемой радиоволны не зависит от направления ее прихода, т.е. $p(\alpha) = p_0$ для всех значений $0 \leq \alpha \leq 2\pi$; при этом общая мощность всех принимаемых сигналов $P_0 = 2\pi p_0$. В этом случае в (8.58) исчезает явная зависимость частоты принимаемого сигнала от направления его прихода и вместо α в качестве аргумента спектральной плотности мощности можно рассматривать теперь независимую переменную f :

$$S(f) = \frac{2G_0 p_0}{\Delta F_{д\max} \sqrt{1 - \left(\frac{f - f_0}{\Delta F_{д\max}}\right)^2}}, \quad |f - f_0| \leq \Delta F_{д\max}, \quad (8.59)$$

и $S(f) = 0$ при $|f - f_0| > \Delta F_{д\max}$.

Таким образом, доплеровский спектр симметричен относительно частоты несущего колебания, излучаемого неподвижным передатчиком, и равен нулю вне полосы частот, характеризующейся граничными частотами $f_0 - F_{д\max}$ и $f_0 + F_{д\max}$, определяемыми максимальными доплеровскими смещениями частоты сигнала, принимаемого движущимся приемником.

График доплеровского спектра (8.59) представлен на рис. 8.8 при $G_0 P_0 = 1$, $\Delta F_{д\max} = 100$ Гц. Особенность этого графика состоит в том, что его значения на граничных частотах $f = f_0 \pm \Delta F_{д\max}$ равны бесконечности. Однако общая мощность сигнала, принимаемого антенной приемника и равная интегралу от спектральной плотности мощности сигнала, конечна.

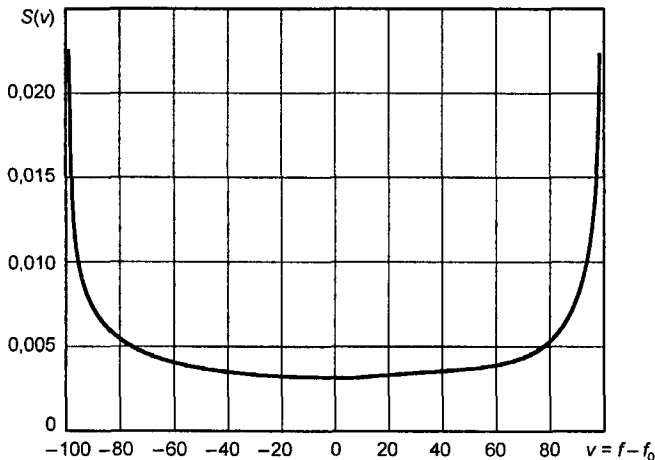


Рис. 8.8. Доплеровский спектр немодулированного несущего колебания ($\Delta F_{д\max} = 100$ Гц)

Доплеровское расширение спектра является следствием многопутевого распространения радиоволн от передатчика к приемнику. Как указывалось ранее, это приводит к замира-

ниям во времени огибающей принимаемого сигнала. Если мощность принимаемого сигнала не зависит от направления прихода волны, а интенсивность городской застройки однородна, то можно принять, что рассмотренная модель описывает процесс формирования одного луча: пути распространения всех радиоволн имеют одинаковую длину, потери распространения для всех волн одинаковы, их мощности на входе приемника одинаковы. Временные характеристики замираний полностью характеризуются формой доплеровского спектра (8.59), имеющего единственный существенный параметр $2\Delta F_{d \max}$ — доплеровское расширение спектра. При такой модели принимаемого поля сигнал в приемнике узкополосный и представляет собой сумму большого числа независимых слагаемых со случайными фазами. Для такого сигнала справедлива центральная предельная теорема — одномерная плотность вероятности мгновенных значений сигнала является гауссовской, а одномерная плотность вероятности мгновенных значений огибающей — релеевской.

Напомним, что этот вывод относится к случаю, когда излучаемый передатчиком сигнал — немодулированное гармоническое колебание. В обычном режиме передачи информации при наличии модуляции доплеровское расширение спектра также имеет место и существенно влияет на способы построения приемных устройств, значительно их усложняя. Тем не менее передача информации и при наличии этого эффекта возможна, а соответствующие методы приема обсуждаются в гл. 9.

8.4.3. Имитационное моделирование многолучевого канала с рассеянием по частоте и по времени

В случае проектирования новых систем связи с подвижными объектами и анализе разных вариантов их построения часто используют имитационные модели сигналов на входе приемника, прошедших многолучевой канал, при имитационном моделировании таких систем. Для получения практически полезных результатов необходимо обеспечить адекватность применяемых моделей сигналов реальным принимаемым сигналам в таких системах. Моделирование сигналов, излучаемых передатчиком, не представляет особых трудностей, поскольку воспроизводятся просто те преобразования, которые реализуются в устройствах передатчика (кодирование, формирование спектра, модуляция и т.д.). Более сложной задачей оказывается моделирование тех преобразований сигнала передатчика, которым он подвергается в канале передачи (в радиоканале, при распространении радиоволн, при отражениях и т.д.). Свойства сигналов при этих преобразованиях обычно исследуются экспериментально путем статистической обработки сигналов на входе приемника при специальных тестовых сигналах, излучаемых передатчиком. Поскольку сигналы узкополосные, то исследования проще проводить посредством статистической обработки квадратурных компонент принимаемых сигналов, которые оказываются гауссовскими процессами со спектральной плотностью мощности, сосредоточенной в окрестности нулевой частоты.

Имитационная модель сигнала на входе приемника должна воспроизводить спектральные и временные характеристики, наиболее близкие к аналогичным характеристикам, получаемым в результате тестовых измерений. Из (8.16) можно получить импульсную характеристику простейшего однолучевого канала с гладкими замираниями, приняв, что $n = 1$. В этом случае комплексная огибающая сигнала на входе приемника представляется просто произведением мгновенных значений комплексной огибающей сигнала $\dot{u}(t)$ на входе канала и коэффициента передачи $\dot{k}(t)$ канала:

$$\dot{v}(t) = \dot{u}(t)\dot{k}(t). \quad (8.60)$$

Поскольку коэффициент передачи $\dot{k}(t) = k_d(t) + jk_m(t)$ является комплексным случайным процессом, то экспериментально изучаются свойства двух вещественных процессов — действительной $k_d(t)$ и мнимой $k_m(t)$ частей этого процесса. В частности, если тестовый сигнал передатчика имеет комплексную огибающую, тождественно равную единице, то реализации комплексной огибающей сигнала на входе приемника с точностью до постоянного множителя совпадают с реализациями коэффициента передачи канала.

Для каналов с гладкими замираниями многочисленные результаты статистической обработки реализаций процессов $k_d(t)$ и $k_m(t)$ позволяют принять следующие их вероятностные модели: независимые гауссовские случайные процессы с равными дисперсиями и одинаковыми спектральными плотностями мощности, форма которых совпадает с формой спектральной плотности мощности принимаемого сигнала, но сосредоточенных около нулевой частоты. В частности, если в канале имеет место доплеровское рассеяние мощности сигнала по частоте, то спектральная плотность принимаемого сигнала является доплеровским спектром, в результате квадратурные компоненты $k_d(t)$ и $k_m(t)$ будут иметь спектральные плотности мощности, соответствующие этому доплеровскому спектру.

Результаты позволяют предложить способ имитационного моделирования радиоканала с аналогичными свойствами с помощью функциональной схемы, представленной на рис. 8.9. Здесь с помощью двух независимых генераторов формируются реализации двух случайных широкополосных гауссовских процессов, каждый из которых имеет равное нулю математическое ожидание и равномерную спектральную плотность мощности в полосе частот, значительно превышающую значение доплеровского расширения спектра. Далее с помощью низкочастотных доплеровских фильтров из этих процессов формируются процессы $k_d(t)$ и $k_m(t)$ с требуемой формой доплеровского спектра.



Рис. 8.9. Имитатор канала с гладкими замираниями

На основе функциональной схемы на рис. 8.9 может быть построена простая цифровая имитационная модель канала с гладкими замираниями, функциональная схема которой представлена на рис. 8.10 [8.3]. Здесь используется генератор комплексных гауссовских случайных чисел для формирования реализаций линейчатых спектров в полосе частот от 0 до $\Delta F_{д \max}$; все числа независимы, имеют равное нулю математическое ожидание и дисперсию, равную единице. Для формирования вещественных реализаций процессов $k_d(t)$ и $k_m(t)$ спектральные компоненты с отрицательными частотами в полосе от $-\Delta f_{д \max}$ до 0, получают как комплексно-сопряженные с соответствующими спектральными компонентами с положительными частотами. Обратное дискретное преобразование Фурье от этих реализаций линейчатых комплексных спектров приводит к вещественным реализациям процессов $k_d(t)$ и $k_m(t)$ во временной области, которые используются в квадратурных ветвях имитатора, представленного на рис. 8.10.

Для получения необходимой формы спектральной плотности мощности этих процессов каждая случайная спектральная компонента умножаются на число $\sqrt{S_d(f)}$, определяемое

значением доплеровского спектра на этой частоте. Для выбора значений доплеровского спектра на границах полосы доплеровского расширения, где функция (8.59) принимает бесконечно большие значения, можно поступить следующим образом: в предпоследних точках на оси частот, ближайших к частотам $\pm \Delta F_{д \max}$, вычисляется наклон функции (8.59), с помощью которого находятся значения усеченного доплеровского спектра на границах полосы.

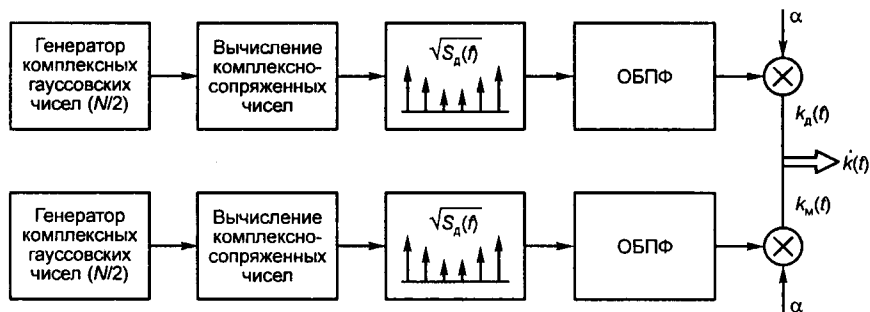


Рис. 8.10. Цифровой имитатор релейского канала с гладкими замираниями

Моделирование с использованием функциональной схемы, представленной на рис. 8.10, реализуется в частотной области на основе комплексного дискретного спектра, длины линий в котором — случайные гауссовские величины. Вещественные реализации процессов $k_d(t)$ и $k_m(t)$ представляют собой ряды Фурье с частотами, эквидистантно расположенными на оси частот в полосе от $-\Delta F_{д \max}$ до $+\Delta F_{д \max}$; коэффициенты этих рядов — комплексно-сопряженные случайные величины. Число дискретных частот N выбирается кратным степени 2 для обеспечения возможности использования быстрых алгоритмов преобразования Фурье. Расстояние между соседними частотами $\Delta f = 2\Delta F_{д \max} / (N - 1)$, что позволяет моделировать реализации комплексного коэффициента передачи на интервале времени $T = 1/(\Delta f)$. Число комплексных случайных чисел, необходимое для моделирования одной реализации процесса, равно N . Число отсчетов этой реализации во временной области равно N , а интервал времени между соседними отсчетами $\Delta t = T/(N - 1) = 1/(\Delta f(N - 1))$.

При достаточно большом значении N можно принять, что процессы $k_d(t)$ и $k_m(t)$ являются независимыми гауссовскими с математическими ожиданиями, равными нулю. Ковариационные функции таких процессов полностью определяются формой доплеровского спектра. Можно показать, что одномерная плотность вероятности модуля $|\dot{k}(t)|$ коэффициента передачи этой модели канала является релейской с параметром

$$\sigma^2 = \sum_{i=-N/2}^{N/2} S_d(f_i). \quad (8.61)$$

Более удобно параметры имитатора выбрать таким образом, чтобы дисперсии процессов $k_d(t)$ и $k_m(t)$ были равны единице. В этом случае параметр (8.61) также должен быть равен единице. Требуемое в каждом частном эксперименте значение параметра σ^2 , определяющее среднее значение уровня сигнала на выходе канала (на входе приемника) можно получить умножением сформированных реализаций этих процессов на коэффициент $\alpha = \sigma$.

Функциональная схема имитатора, представленная на рис. 8.10, обеспечивает формирование реализаций комплексного коэффициента передачи однолучевого релейского канала.

Этот же принцип моделирования можно положить в основу построения модели многолучевого канала (8.16) с гладкими релейскими независимыми замираниями каждого луча (рис. 8.11). Каждый луч с номером i формируется аналогичным имитатором, а лучи отличаются друг от друга только относительным временем задержки τ_i и интенсивностью α_i .

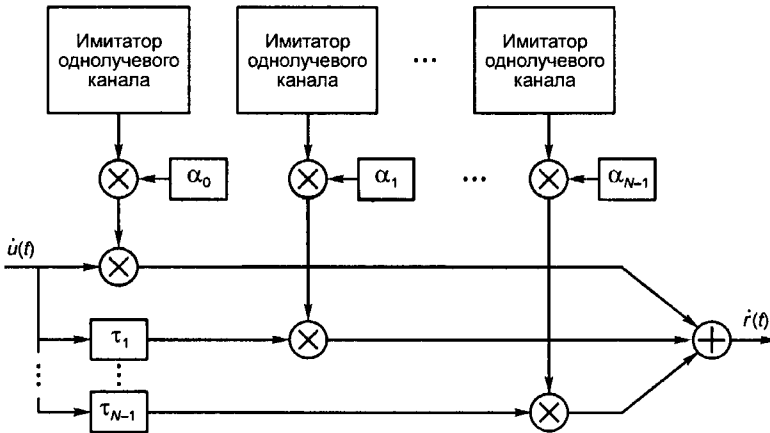


Рис. 8.11. Цифровой имитатор многолучевого релейского канала с гладкими замираниями

Если одна из спектральных компонент $\sqrt{S_d(f_i)}$, $i = 1, 2, \dots, N$ имеет заметно большее значение по сравнению с остальными, а соответствующий этой частоте коэффициент ряда Фурье отличен от нуля и не является случайной величиной, то распределение огибающей соответствующего луча изменяется, переходя от распределения Релея к распределению Райса. Данный факт можно использовать для изменения распределений огибающих отдельных лучей многолучевого канала. Если один из имитаторов однолучевого канала заменить генератором константы, то соответствующий канал можно рассматривать как имитатор прямого луча, не подверженного замираниям.

8.5. Модели каналов для испытаний оборудования некоторых радиосистем

8.5.1. Рекомендации стандарта GSM-900 для наземных сетей

Радиоинтерфейс систем связи с подвижными объектами должен обеспечивать достаточно высокое качество передачи данных в довольно сложных условиях распространения радиоволн: в интенсивной городской застройке, как правило, при отсутствии прямой видимости между антеннами передатчика и приемника, при достаточно быстром перемещении подвижной станции. Вместе с тем не исключены случаи, когда передатчик и приемник находятся достаточно близко друг от друга, имеется прямая видимость между их антеннами, а подвижная станция малоподвижна, т.е. диапазон изменения условий распространения радиоволн для таких радиосистем достаточно широк. Для контроля качества передачи информации в таких условиях стандартом GSM-900 рекомендуется несколько испытательных

моделей радиоканала, которые могут быть использованы при моделировании основных элементов радиоинтерфейса на этапе их разработки либо при испытаниях соответствующего оборудования на специальных стендах.

Оборудование радиоинтерфейса стандарта GSM-900 рассчитывается на работу в радиоканалах, обладающих явно выраженными дисперсионными свойствами, где имеют место существенные многолучевость, расширение задержки и доплеровское расширение спектра. Каждый путь распространения радиоволны между базовой и подвижной станциями можно представить себе как ломанную линию, содержащую большое число отражателей и рассеивателей, расположенных случайным образом в окрестности подвижной станции. Для каждого такого пути амплитуда A принимаемого сигнала, его временная задержка τ относительно сигнала, пришедшего по кратчайшему пути, и доплеровское смещение F частоты несущего колебания принимают некоторые частные значения. Таких путей в принципе может быть неограниченно много. Поэтому для комплексной огибающей принимаемого сигнала в общем случае приходится использовать следующее представление:

$$\dot{b}(t) = \iint_{R^2} \dot{a}(t - \tau) \exp\{j2\pi F(t - \tau)\} A(\tau, F) d\tau dF,$$

где $\dot{a}(t - \tau)$ — комплексная огибающая задержанного сигнала; $A(\tau, F)$ — множитель, определяющий изменение амплитуды сигнала с задержкой τ и доплеровским смещением частоты F , характеризует рассеивающие свойства радиоканала. Функция $A(\tau, F)$ может существенно изменяться при изменении положения подвижной станции. Полезно иметь среднее значение $\overline{A^2(\tau, F)}$ квадрата этой функции для малой окрестности подвижной станции, ограниченной, например, кругом радиусом 10 м, для которой можно считать справедливым предположение о стационарности в широком смысле и слабой коррелированности значений данной функции. Сечение последней вдоль оси F при фиксированном значении $\tau = \tau'$ представляет собой доплеровское расширение спектра $S(f) = \overline{A^2(\tau', F)}$; данную функцию называют *доплеровским спектром*. Сечение функции $\overline{A^2(\tau, F)}$ вдоль оси τ при фиксированном значении $F = F'$ имеют *профилем задержек* $H(\tau) = \overline{A^2(\tau, F')}$. Функция $S(f)$ характеризует усредненное рассеяние радиоканалом мощности принимаемого сигнала по частоте, а функция $H(\tau)$ — по времени. Усредненные функции используются для моделирования радиоканала.

Для практического моделирования радиоканала вместо функции $\overline{A^2(\tau, F)}$, описывающей непрерывное рассеяние мощности принимаемого сигнала по частоте и по времени, рекомендуются модели с дискретной многолучевостью, которые строятся на основе линий задержек с отводами. Каждый отвод соответствует одной лучевой компоненте; для каждого отвода устанавливаются значения времени задержки относительно первого луча и средней мощности соответствующей лучевой компоненты; огибающая сигнала каждого луча случайным образом изменяется во времени и имеет заданное одномерное распределение; скорость изменений огибающей во времени связана с формой доплеровского спектра этой компоненты.

Стандартом GSM-900 определяются два типа доплеровского спектра, рекомендуемые для моделирования радиоканала:

а) *КЛАСС* — классический доплеровский спектр

$$S(F) = A/\left[\sqrt{1 - (F/\Delta F_{\text{д макс}})^2}\right] \text{ для } F \in (-\Delta F_{\text{д макс}}, \Delta F_{\text{д макс}}), \quad (8.62)$$

который используется во всех случаях (см. (8.59));

б) РАЙС — сумма классического доплеровского спектра и спектра сигнала, пришедшего к приемнику по кратчайшему пути; принимается, что общая мощность многолучевых компонент равна мощности сигнала, пришедшего по кратчайшему пути:

$$S(f) = 0,41 / \left(2\pi\Delta F_{\text{д max}} \sqrt{1 - (F / \Delta F_{\text{д max}})^2} \right) + 0,918(F - 0,7\Delta F_{\text{д max}}), \quad (8.63)$$

для $F \in (-\Delta F_{\text{д max}}, \Delta F_{\text{д max}})$.

Для адекватного представления типичных условий распространения принято считать достаточными модели с 12 лучами. Однако если нет возможности использовать такие полные модели, для облегчения задач моделирования допускается применение упрощенных моделей, содержащих всего шесть лучевых компонент. Эти упрощенные модели могут быть применены, в частности, для моделирования многолучевых мешающих сигналов. Всякий раз, когда это оказывается возможным, рекомендуется использовать полную модель. Для каждой модели стандартом определены две альтернативы, обозначенные в приводимых ниже таблицах цифрами (1) и (2) в соответствующих столбцах. Разные лучевые компоненты могут иметь разные типы доплеровского спектра и разные значения параметра $\Delta F_{\text{д max}} = v/\lambda$, где v — скорость перемещения подвижной станции, м/с; λ — длина волны, м; эта возможность может быть использована, если необходимо учесть направление перемещения подвижной станции.

В табл. 8.2–8.8 приведены численные значения параметров многолучевых компонент, рекомендуемые стандартом GSM-900 [8.4].

Таблица 8.2. Типичный случай сельской местности (6 лучей)

Номер луча	Время задержки относительно первого луча, мкс		Среднее относительное значение мощности, дБ		Тип доплеровского спектра
	(1)	(2)	(1)	(2)	
1	0,0	0,0	0,0	0,0	РАЙС
2	0,1	0,2	-4,0	-2,0	КЛАСС
3	0,2	0,4	-8,0	-10,0	То же
4	0,3	0,6	-12,0	-20,0	- » -
5	0,4	-	-16,0	-	- » -
6	0,5	-	-20,0	-	- » -

Таблица 8.3. Типичный случай холмистой местности (12 лучей)

Номер луча	Время задержки относительно первого луча, мкс		Среднее относительное значение мощности, дБ		Тип доплеровского спектра
	(1)	(2)	(1)	(2)	
1	0,0	0,0	-10,0	-10,0	КЛАСС
2	0,1	0,2	-8,0	-8,0	То же
3	0,3	0,4	-6,0	-6,0	- » -
4	0,5	0,6	-4,0	-4,0	- » -
5	0,7	0,8	0,0	0,0	- » -
6	1,0	2,0	0,0	0,0	- » -
7	1,3	2,4	-4,0	-4,0	- » -
8	15,0	15,0	-8,0	-8,0	- » -
9	15,2	15,2	-9,0	-9,0	- » -
10	15,7	15,8	-10,0	-10,0	- » -
11	17,2	17,2	-12,0	-12,0	- » -
12	20,0	20,0	-14,0	-14,0	- » -

Таблица 8.4. Типичный случай холмистой местности (упрощенный вариант, 6 лучей)

Номер луча	Время задержки относительно первого луча, мкс		Среднее относительное значение мощности, дБ		Тип доплеровского спектра
	(1)	(2)	(1)	(2)	
1	0,0	0,0	0,0	0,0	КЛАСС
2	0,1	0,2	-1,5	-2,0	То же
3	0,3	0,4	-4,5	-4,0	- » -
4	0,5	0,6	-7,5	-7,0	- » -
5	15,0	15,0	-8,0	-6,0	- » -
6	17,2	17,2	-17,7	-12,0	- » -

Таблица 8.5. Типичный случай городской застройки (12 лучей)

Номер луча	Время задержки относительно первого луча, мкс		Среднее относительное значение мощности, дБ		Тип доплеровского спектра
	(1)	(2)	(1)	(2)	
1	0,0	0,0	-10,0	-10,0	КЛАСС
2	0,1	0,2	-8,0	-8,0	То же
3	0,3	0,4	-6,0	-6,0	- » -
4	0,5	0,6	-4,0	-4,0	- » -
5	0,7	0,8	0,0	0,0	- » -
6	1,0	2,0	0,0	0,0	- » -
7	1,3	2,4	-4,0	-4,0	- » -
8	15,0	15,0	-8,0	-8,0	- » -
9	15,2	15,2	-9,0	-9,0	- » -
10	15,7	15,8	-10,0	-10,0	- » -
11	17,2	17,2	-12,0	-12,0	- » -
12	20,0	20,0	-14,0	-14,0	- » -

Таблица 8.6. Типичный случай городской застройки (упрощенный вариант, 6 лучей)

Номер луча	Время задержки относительно первого луча, мкс		Среднее относительное значение мощности, дБ		Тип доплеровского спектра
	(1)	(2)	(1)	(2)	
1	0,0	0,0	-3,0	-3,0	КЛАСС
2	0,2	0,2	0,0	0,0	То же
3	0,5	0,6	-2,0	-2,0	- » -
4	1,6	1,6	-6,0	-6,0	- » -
5	2,3	2,4	-8,0	-8,0	- » -
6	5,0	5,0	-10,0	-10,0	- » -

Таблица 8.7. Профиль для контроля эквалайзера (6 лучей)

Номер луча	Время задержки относительно первого луча, мкс	Среднее относительное значение мощности, дБ	Тип доплеровского спектра
1	0,0	0,0	КЛАСС
2	3,2	0,0	То же
3	6,4	0,0	- » -
4	9,6	0,0	- » -
5	12,8	0,0	- » -
6	16,0	0,0	- » -

Таблица 8.8. Типичный случай для очень малых сот (2 луча)

Номер луча	Время задержки относительно первого луча, мкс	Среднее относительное значение мощности, дБ	Тип доплеровского спектра
1	0,0	0,0	КЛАСС
2	0,4	0,0	То же

В 8.4.3 приведены сведения о функциональной схеме имитатора многолучевого канала с дискретной многолучевостью, численные значения параметров которого могут быть взяты в соответствии с данными приведенных выше таблиц.

8.5.2. Рекомендации стандарта GMR для спутниковых систем

Условия распространения сигнала в спутниковом радиоканале подвижной связи также являются достаточно сложными, поскольку существует много факторов, существенно влияющих на принимаемый сигнал. Среди них наиболее важными признаются эффект рассеяния электромагнитной волны в процессе распространения (потери распространения в свободном пространстве или условия прямой видимости), погодные условия и характер земной поверхности в окрестности спутникового терминала, изменяющийся при перемещении терминала и спутника. Для таких условий распространения предложено несколько моделей спутникового радиоканала.

Статистические модели с одним состоянием. В свободном пространстве мощность сигнала, принимаемого приемником, находящимся на расстоянии d от передатчика, определяется соотношением

$$P_{\text{пр}}(d) = P_{\text{зи}} \frac{S_{\text{пр}}}{4\pi d^2},$$

где $P_{\text{зи}}$ — эффективно излучаемая передатчиком мощность; $S_{\text{пр}}$ — эффективная площадь апертуры приемной антенны. Потери распространения

$$\Pi_d = \frac{P_{\text{пр}}}{P_{\text{зи}}} = \frac{S_{\text{пр}}}{4\pi d^2} \tag{8.64}$$

обусловлены увеличением площади эквивалентной сферы радиусом d . Можно принять, что при перемещении приемника в некоторой локальной зоне по поверхности Земли расстояние d остается практически неизменным; значение потерь Π_d при этом остается неизменным.

При распространении электромагнитной волны от спутника к наземному терминалу часть энергии волны поглощается атмосферой Земли. Статистическая обработка многочисленных результатов измерений потерь распространения в тропосфере при угле склонения не менее 14° обеспечивает возможность выбора адекватной вероятностной модели для коэффициента ослабления $\alpha_{\text{тр}}$ в виде гауссовской случайной величины с плотностью вероятности [8.5]

$$w_\alpha(a) = \frac{1}{\sqrt{2\pi}\sigma_\alpha} \exp\left[-(a - m_\alpha)^2 / (2\sigma_\alpha^2)\right]. \tag{8.65}$$

Особенно заметный вклад в потери распространения вносят облака и дожди, которые могут значительно ухудшить качество передачи при грозах и ураганах. Для потерь в дожде также накоплен большой экспериментальный статистический материал, в соответствии с которым огибающую V сигнала можно рассматривать как случайную величину с плотностью вероятности [8.5]

$$w_\nu(\nu) = \frac{1}{\sqrt{2\pi}\sigma_\nu\nu} \exp\left[-(\ln \nu - m_\nu)^2 / (2\sigma_\nu^2)\right], \nu \geq 0 \quad (8.66)$$

логарифмически нормального распределения. Потери в дожде существенно зависят от частоты электромагнитной волны, возрастая с повышением частоты. Можно принять, что при частоте несущего колебания меньше 10 ГГц потери в дожде малы и ими можно пренебречь. Если же частота волны больше 10 ГГц, то потери следует учитывать.

Значительное влияние на условия приема оказывают предметы, находящиеся в окрестности приемника на земной поверхности. Способы описания этого влияния оказываются наиболее сложными. Наиболее часто используемая модель для принимаемого сигнала представляет собой сумму двух компонент: сигнал, пришедший по прямому пути от спутника к антенне приемника, и рассеянная многолучевая компонента. Амплитуду V прямого сигнала рекомендуется рассматривать как случайную величину с логарифмически нормальным распределением со средним значением m_ν и стандартным отклонением σ_ν . Одномерное распределение огибающей R многолучевой компоненты принимается релейевским с единственным параметром σ_R^2 , характеризующим среднюю мощность этой компоненты. Условное одномерное распределение огибающей Z суммы прямого сигнала и многолучевой компоненты при фиксированном значении ν огибающей V прямого сигнала будет иметь плотность Релея–Райса [8.5]

$$w(z|\nu) = \frac{z}{\sigma_R^2} \exp\left[-\frac{z^2 + \nu^2}{2\sigma_R^2}\right] I_0\left(\frac{z\nu}{\sigma_R^2}\right). \quad (8.67)$$

Если учесть, что огибающая прямого луча с учетом дождя представляет собой логарифмически гауссовскую случайную величину с плотностью (8.66), то можно найти безусловную плотность для одномерного распределения огибающей сигнала, принимаемого наземным терминалом:

$$w_z(z) = \frac{z}{\sigma_R^2 \sqrt{2\pi}\sigma_\nu} \int_0^\infty \frac{1}{\nu} \exp\left[-\frac{(\ln \nu - m_\nu)^2}{2\sigma_\nu^2} - \frac{z^2 + \nu^2}{2\sigma_R^2}\right] I_0\left(\frac{z\nu}{\sigma_R^2}\right) d\nu. \quad (8.68)$$

Численные значения параметров приведенных здесь распределений зависят от положения спутника на небосклоне. При большом угле возвышения вклад прямого сигнала в суммарную мощность принимаемого сигнала является определяющим; влияние предметов, окружающих спутниковый терминал, незначительно, в результате чего многолучевая компонента вносит лишь незначительную долю в суммарную мощность принимаемого сигнала. При уменьшении угла возвышения существенно возрастает влияние окружающих терминал предметов (зданий, деревьев и т.д.) и рельефа местности в окрестности терминала; в результате многолучевая компонента принимаемого сигнала может оказаться определяющей. Поэтому угол возвышения является параметром выбираемой модели спутникового канала.

Каждое из приведенных выше распределений (8.65)–(8.68) представляет собой модель потерь распространения из-за одного физического эффекта — поглощение в атмосфере Земли, наличие дождя, влияние предметов, окружающих спутниковый терминал. Плотность распределения (8.68) описывает совместное влияние сразу двух факторов — дождя и местных предметов. Кроме того, эти распределения описывают лишь вероятностные свойства огибающей принимаемого сигнала в один единственный момент времени. Такие модели оказываются полезными при вычислении средней вероятности ошибки при передаче цифровых данных без помехоустойчивого кодирования либо при использовании такого кодирования и идеального перемежения, при котором ошибки можно считать независимыми, при вычислении вероятности отказа, расчете допустимых значений энергетических параметров радиолинии и т.д.

Рекомендации стандарта GMR. Модель канала с райсовскими замираниями рекомендуется использовать для моделирования быстрых замираний, т.е. без учета потерь в атмосфере Земли и затенений. Модель принимаемого сигнала представляет собой сумму двух компонент: прямого сигнала, принятого по линии прямой видимости между передатчиком и приемником, и многолучевой диффузионной компоненты [8.6]. Прямой сигнал имеет не изменяющуюся во времени среднюю мощность и постоянное доплеровское смещение $F_{дпр}$ частоты несущего колебания (относительно частоты неподвижного передатчика). Многолучевая компонента является суммой большого числа отраженных сигналов, приходящих к приемной антенне по разным путям с азимутальными углами, распределенными равномерно в интервале 0 до 2π . Возможные значения доплеровских смещений частоты таких сигналов заполняют интервал $[-f_{д\max}, f_{д\max}]$, где $f_{д\max}$ — максимальное смещение частоты, имеющее место при движении приемника и передатчика навстречу друг другу. Поэтому многолучевая компонента может рассматриваться как узкополосный гауссовский стационарный случайный процесс с нулевым математическим ожиданием и спектральной плотностью мощности (8.59). Одномерная плотность вероятности такого процесса является релейской с единственным параметром σ^2 , определяющим мощность этого процесса. Огибающая суммы прямого сигнала и такого гауссовского процесса будет иметь одномерную плотность вероятности Райса, для которой удобно ввести параметр

$$K = \frac{P_{прс}}{P_R} = \frac{\nu^2}{2\sigma^2}, \tag{8.69}$$

называемый коэффициентом Райса, который определяется отношением мощности $P_{прс}$ прямого сигнала к мощности $P_R = \sigma^2$ многолучевой релейской компоненты. Полезная информация содержится только в прямом сигнале; только мощность этого сигнала должна учитываться при определении значения параметра спутниковой линии E_b/N_0 — отношения энергии сигнала на 1 бит передаваемой информации к спектральной плотности мощности шума.

Плотность (8.67) содержит два параметра: ν — амплитуду прямого сигнала; σ^2 — мощность релейской компоненты. Второй начальный момент этого распределения $M[Z^2] = \nu^2 + 2\sigma^2$, а нормированная случайная величина $U = Z / \sqrt{\nu^2 + 2\sigma^2}$ имеет плотность вероятности с одним параметром K :

$$w(u) = 2(K + 1)u \exp\{-u^2(K + 1) - K\} I_0\left(2u\sqrt{K(K + 1)}\right). \tag{8.70}$$

Для линии со спутником, находящимся на геостационарной орбите, необходимо учитывать движение наземного терминала. В этом случае различные лучевые компоненты с разными углами прихода по отношению к направлению движения терминала, будут иметь различные значения доплеровского смещения частоты. Если можно принять условия, для которых была построена модель Кларка (см. 8.4.1), то доплеровский спектр будет определяться равенством (8.59). Ковариационная функция, соответствующая такому спектру имеет вид

$$R(\tau) = \sigma^2 J_0(2\pi\Delta F_{д\max}\tau). \tag{8.71}$$

Если же спутник находится не на геостационарной орбите, то необходимо учитывать дополнительно доплеровское смещение частоты принимаемого сигнала, возникающее из-за движения спутника с относительной скоростью $v_{сп}$: $\Delta F_{дсп} = f_0 v_{сп}/c$; такое смещение будут испытывать все спектральные компоненты спектра (8.59), который теперь будет представлять только непрерывную компоненту доплеровского спектра. Если присутствует прямой сигнал с амплитудой ν , то результирующий низкочастотный доплеровский спектр, кроме непрерывной компоненты, будет содержать дискретную компоненту

$$S(F) = \frac{\sigma^2}{\pi \Delta F_{\text{д max}} \sqrt{1 - \left(\frac{F - \Delta F_{\text{д сп}}}{\Delta F_{\text{д max}}} \right)^2}} + \frac{\nu^2}{2} \delta(F - \Delta F_{\text{д сп}} - F_{\text{д пр с}}), \quad |F - \Delta F_{\text{д сп}}| \leq \Delta F_{\text{д max}}. \quad (8.72)$$

На рис. 8.12 представлена структурная схема имитатора одной лучевой компоненты спутникового канала с райсовскими замираниями, рекомендуемая стандартом GMR.

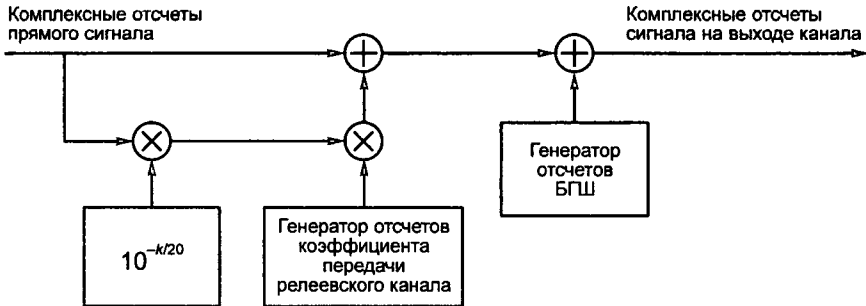


Рис. 8.12. Функциональная схема имитатора спутникового канала

Отсчет комплексной огибающей прямого сигнала умножается сначала на постоянный множитель $10^{-K/20}$ (K , дБ), а затем на комплексный отсчет значений узкополосного процесса, имитирующего многолучевую компоненту с релейским распределением огибающей; полученные отсчеты складываются, образуя комплексный отсчет сигнала на выходе спутникового канала с райсовскими замираниями. Наконец, к полученному отсчету сигнала добавляется комплексный отсчет БГШ, предварительно нормированный таким образом, чтобы обеспечить требуемое значение отношения E_b/N_0 .

Если ширина спектра сигнала оказывается больше полосы когерентности канала, то возникают частотно-селективные замирания. Дискретизируя время задержек многолучевых компонент, непрерывный импульсный отклик канала можно приближенно представить следующим выражением:

$$h(t, \tau) = \sum_{i=0}^{n-1} k_i(t) \delta(\tau - \tau_i(t)) \exp\{j[2\pi F_{d,i} t + \theta_i(t)]\}, \quad (8.73)$$

где $k_i(t)$, $\tau_i(t)$, $F_{d,i}$, $\theta_i(t)$ — параметры i -го луча: амплитуда, задержка, доплеровское смещение частоты и фаза. Эта модель канала как и ранее представляется линией задержки с отводами и соответствующими весовыми коэффициентами. Здесь лучевая компонента с индексом $i = 0$ соответствует прямому сигналу. Каждый луч может описываться одной из узкополосных моделей, представленных выше. Все лучи обычно предполагаются статистически независимыми.

Лучевые компоненты часто разделяют на три группы с различными свойствами лучей: прямой луч, близкие отражения, дальние отражения. Прямой луч моделируется с использованием марковской цепи с двумя состояниями (хорошее и плохое), соответствующими отсутствию и наличию затенений. Близкие отраженные сигналы имеют небольшие временные задержки относительно прямого сигнала, не превышающие некоторого значения $\tau_{бл}$, зависящего от конкретной обстановки в окрестности приемника и от угла возвышения спутника и составляющего несколько сотен наносекунд. Дальние отражения имеют задержки порядка 10–15 мкс.

Численные значения параметров моделей и типы замираний отдельных лучевых компонент, рекомендуемые стандартом GMR, приведены в табл. 8.9, 8.10.

Таблица 8.9. Модели прямого луча прямой линии спутникового канала

К, дБ	Номер луча	Задержка	Мощность	Тип замираний
–	1	0	$P_{пр с}$	Прямой сигнал
–	1	0	$P_{рел}$	Релей
9	1	0	$P_{пр с}$	Прямой сигнал
	1	0	$P_{пр с}/8$	Релей
6	1	0	$P_{пр с}$	Прямой сигнал
	1	0	$P_{пр с}/4$	Релей

Таблица 8.10. Модели спутникового канала прямой линии с райсовскими замираниями

Параметр Райса К, дБ	Номер луча	Задержка D = 12 мкс	Мощность	Тип замираний
6	1	0	$P_{пр с}$	Прямой сигнал
	1	0	$P_{пр с}/20$	Релей
	2	D/4	$P_{пр с}/20$	То же
	3	D/2	$P_{пр с}/20$	– » –
	4	3D/4	$P_{пр с}/20$	– » –
	5	D	$P_{пр с}/20$	– » –

Важным параметром замираний принимаемого сигнала признается среднее значение T_X длительности интервала времени, в течение которого огибающая сигнала имеет значение, меньшее заданного уровня X [8.7]:

$$T_X = \frac{P\{r(t) \leq X\}}{N_X} \cong \frac{\sum_i \tau_i}{TN_X}, \tag{8.74}$$

где τ_i — длительность i -го интервала занижения на интервале наблюдения T и

$$N_X = \int_0^\infty r' w(X, r') dr' \tag{8.75}$$

средняя скорость пересечений уровня X огибающей сигнала с положительной производной; r' — производная по времени огибающей $r(t)$; $w(r, r')$ — совместная плотность вероятности огибающей и ее производной в момент времени t . Для доплеровского спектра (8.59) из (8.75) можно получить [8.7]:

$$T_X = \frac{\exp(\bar{X}^2) - 1}{\bar{X} F_{д \max} \sqrt{2\pi}}, \tag{8.76}$$

где $\bar{X} = X / (\sigma\sqrt{2})$. Параметры (8.75) и (8.76) могут быть использованы, например, при выборе оптимального значения интервала перемежения кодовых символов в передатчике, что позволяет уменьшить корреляцию между ошибками в канале.

Статистические модели с несколькими состояниями. В общем случае при выборе модели спутникового канала приходится учитывать все перечисленные выше воздействия на принимаемый сигнал. Потери, обусловленные перечисленными факторами, непрерывно меняются во времени. Например, подвижная станция перемещается из городской застрой-

ки, где прямой сигнал отсутствовал (плохое состояние канала), в пригород с другими условиями распространения радиоволн (появляется прямой сигнал, хорошее состояние канала). Безусловное распределение огибающей принимаемого сигнала с плотностью (8.68) слишком сложно для практических вычислений и построения соответствующих моделей. Более удобной для практического использования считается модель в виде цепи Маркова с несколькими состояниями, каждое из которых соответствует определенному состоянию спутникового канала [8.7]. Важными при выборе модели спутникового канала являются следующие факторы: тропосфера, облака, дождь, затеняющие деревья, прерывающие путь распространения волн здания. Последствия влияния элементов каждой группы можно описать лишь статистически, используя вероятностные модели в виде случайных величин для параметров распространяющегося сигнала. Случайные величины, выбранные для разных групп, можно принять независимыми. Например, представляется вполне допустимым предположение, что потери в атмосфере Земли и потери, обусловленные особенностями рельефа и застройки окрестности спутникового терминала, независимы. В соответствии с этим предположением модель канала в виде цепи Маркова можно расщепить на два слоя: слой пространственной цепи и слой наземной цепи. В слое пространственной цепи рассчитываются потери распространения в космическом пространстве, в тропосфере и в атмосфере Земли; в слое наземной цепи прогнозируются потери из-за особенностей рельефа местности и застройки земной поверхности. Такое разделение всех факторов влияния на две группы, позволяет уменьшить число возможных состояний цепи Маркова и снизить сложность вычислений при практическом моделировании.

Пространственные цепи Маркова. Среди пространственных факторов наиболее значимыми являются погодные условия, которые существенно различны для различных климатических поясов. Погодные условия можно разделить на четыре подгруппы: нет дождя, прерывистый дождь, дождь, гроза (шторм). Для этих подгрупп можно ввести следующие состояния пространственной цепи Маркова [8.8]:

- SS1 — нет дождя;
- SS2 — дождь;
- SS3 — прерывистый дождь;
- SS4 — гроза.

Любая стационарная цепь Маркова полностью определяется двумя матрицами: \mathbf{W}_s — матрица (вектор) вероятностей состояний и \mathbf{P}_s — матрица вероятностей переходов. Элемент $P_{s_{ij}}$ матрицы \mathbf{P}_s представляет собой вероятность перехода из состояния i в состояние j . Безусловная (полная) вероятность состояния i определяется i -м элементом матрицы \mathbf{W}_s . Численные значения этих вероятностей можно оценить путем статистической обработки результатов наблюдений. Например, можно принять, что

$$\mathbf{W}_s = N_i / N, \quad \mathbf{P}_s = N_{ij} / N, \quad (8.77)$$

где N — общее число проведенных наблюдений, длительность каждого из которых была принята одинаковой и равной T_0 ; N_i — число наблюдений, в которых имело место состояние i ; N_{ij} — число наблюдений, в которых имел место переход из состояния i в состояние j . Для получения достаточно точных оценок элементов матриц \mathbf{W}_s и \mathbf{P}_s с помощью соотношений (8.77) приходится выбирать общее число наблюдений N столь большим, чтобы получить достаточно большие значения N_i и N_{ij} .

Указанные выше матрицы вероятностей для марковских цепей обладают следующими свойствами:

$$\sum_{i=1}^4 \mathbf{W}_s = 1; \quad \sum_{j=1}^4 \mathbf{P}_s = 1, \quad i = 1, 2, 3, 4; \quad \mathbf{W}_s^T \cdot \mathbf{P}_s = \mathbf{W}_s.$$

Наземная цепь Маркова. Было рекомендовано использовать три типа условий приема, каждому из которых целесообразно определить соответствующее состояние цепи [8.9], [8.10]:

- ST1 — условия прямой видимости;
- ST2 — условия умеренного затенения;
- ST3 — условия сильного затенения.

Структуры матрицы вероятностей состояний W_s и матрицы вероятностей переходов P_s такие же, как и для пространственных цепей Маркова. Однако следует отметить, что среднее значение времени, в течение которого состояние цепи сохраняется неизменным, в данном случае оказывается меньше по сравнению с пространственной цепью Маркова.

Модель канала. Потери Π_d в свободном пространстве и $\alpha_{тр}$ в тропосфере существуют всегда. Для потерь из-за дождя вводятся четыре возможных значения, каждому из которых соответствует одно состояний пространственной цепи Маркова и своя пара значений параметров m_ν и σ_ν в (8.68); потери из-за наличия отражений от местных предметов в окрестности приемника характеризуются значением отношения $\nu^2/(2\sigma^2)$ мощности $\nu^2/2$ прямого сигнала к мощности σ^2 многолучевой компоненты; для потерь из-за затенений вводятся три возможных значения, каждому из которых соответствует одно состояние наземной цепи Маркова.

Общий коэффициент ослабления сигнала в спутниковом канале можно представить как произведение коэффициентов ослабления Π_d , рассчитываемом по (8.62), $\Pi_{пр м}$, рассчитываемом с использованием пространственной цепи Маркова, и $\Pi_{наз м}$, рассчитываемом с применением наземной цепи Маркова.

8.5.3. Модели каналов некоторых широкополосных систем третьего поколения

В последние годы были предприняты значительные усилия для выработки общих рекомендаций по построению систем будущих поколений и соответствующего единого глобального стандарта. Большую роль в организации совместной работы национальных и региональных производителей оборудования для таких систем выполняет Международный союз электросвязи (МСЭ).

Модели МСЭ. К настоящему времени разработаны рекомендации по использованию моделей радиоканалов для наземных сотовых систем в виде линий задержек с отводами. Эти модели предусматривают три типа условий распространения: внутри зданий, для медленных абонентов (пешеходов) внутри и вне зданий, для транспортных средств (с высокими антеннами базовых станций). Для каждого типа условий определены две модели: модель *A* для малых расширений задержек и модель *B* для больших расширений задержек (для наилучшего сценария). Параметры этих моделей приведены в табл. 8.11–8.13 [8.7].

Таблица 8.11. Модели МСЭ радиоканалов сотовых систем внутри зданий

Номер луча	Модель A (50%)		Модель B (45%)		Доплеровский спектр
	Задержка, нс	Мощность, дБ	Задержка, нс	Мощность, дБ	
1	0	0	0	0	Равномерный
2	50	-3,0	100	-3,6	То же
3	110	-10,0	200	-7,2	- » -
4	170	-18,0	300	-10,8	- » -
5	290	-26,0	500	-18	- » -
6	310	-32,0	700	-25,2	- » -

В табл. 8.11 указаны также численные значения оценок относительной частоты появления этих моделей в реальных условиях. Принято, что замирания подчиняются распределению Релея, а доплеровский спектр является равномерным в полосе от $-F_{д\max}$ до $+F_{д\max}$. Для этой модели рекомендован и способ прогноза потерь (дБ) распространения на расстоянии d :

$$\Pi(d) = 37 + 30 \lg(d) + 18N_{зт} \left(\frac{N_{п+2}}{N_{п+1}} - 0,46 \right). \quad (8.78)$$

Потери распространения для пешеходов (табл. 8.12) рекомендуется рассчитывать в следующем виде:

$$\Pi(d) = 49 + 40 \lg(d) + 30 \lg(f_0). \quad (8.79)$$

Таблица 8.12. Модели МСЭ радиоканалов сотовых систем для пешеходов

Номер луча	Модель А (50%)		Модель В (45%)		Доплеровский спектр
	Задержка, нс	Мощность, дБ	Задержка, нс	Мощность, дБ	
1	0	0	0	0	Классический
2	110	-9,7	200	-0,9	То же
3	190	-19,2	800	-8,0	- » -
4	410	-22,8	1200	-7,8	- » -
5	-	-	2300	-18	- » -
6	-	-	3700	-23,9	- » -

Если рассматривается случай связи между терминалом вне здания с терминалом внутри здания, то необходимо добавить потери, связанные с проникновением радиоволн в здание, которые рекомендуется рассматривать как гауссовскую случайную величину со средним значением 12 дБ и стандартным отклонением 8 дБ. Доплеровский спектр в полосе от $-F_{д\max}$ до $+F_{д\max}$ предполагается классическим (модель Кларка, см. (8.59)).

Потери распространения для транспортных средств (табл. 8.13) рекомендуется вычислять как

$$\Pi(d) = 80 + 40(1 - 4 \cdot 10^{-3} \Delta h_0) \lg(d) - 18 \lg(\Delta h_0) + 21 \lg(f_0), \quad (8.80)$$

где f_0 — частота несущего колебания; Δh — высота установки антенны базовой станции относительно плоской крыши; формулу рекомендуется использовать при условии $0 < \Delta h_0 < 50$ м.

Таблица 8.13. Модели МСЭ радиоканалов сотовых систем для транспортных средств

Номер луча	Модель А (50%)		Модель В (45%)		Доплеровский спектр
	Задержка, нс	Мощность, дБ	Задержка, нс	Мощность, дБ	
1	0	0	0	-2,5	Классический
2	310	-1	300	0	То же
3	710	-9	8900	-12,8	- » -
4	1090	-10	12 900	-10	- » -
5	1730	-15	17 100	-25,2	- » -
6	2510	-20	20 000	-16	- » -

Во всех трех случаях накладываются случайные потери затенений, для которых рекомендуется модель в виде логарифмически нормальной случайной величины. Рассеяние этих потерь составляет: внутри зданий — 12 дБ; для транспортных средств — 10 дБ; для пешеходов вне зданий — 10 дБ; для пешеходов внутри зданий — 12 дБ. Автокорреляционная функция затенений выбрана экспоненциальной

$$R(x) = \exp\{-\ln(2)|x|/d_{\text{кор}}\},$$

в которой радиус корреляции $d_{кор}$ для транспортных средств принят 20 м (для других условий не определен).

Модели стандартов IEEE 802.11 и HIPERLAN ETSI. В отличие от моделей МСЭ и стандарта GSM, рекомендуемых для сотовых систем, эти модели радиоканалов в виде линий задержек с отводами рекомендуются для использования в других условиях и при значительно более широких полосах радиочастот:

- модель *A* — для типовых офисных условий, при отсутствии прямого сигнала, при среднеквадратическом значении расширения задержки 50 нс;
- модель *B* — для типовых офисных условий и открытого пространства, при отсутствии прямого сигнала, при среднеквадратическом значении расширения задержки 100 нс;
- модель *C* — для большого открытого пространства (внутри и вне помещений), при отсутствии прямого сигнала, при среднеквадратическом значении расширения задержки 150 нс;
- модель *D* — для большого открытого пространства, присутствует прямой сигнал, среднеквадратическое значение расширения задержки 140 нс;
- модель *E* — для типового большого открытого пространства (внутри и вне помещений), при отсутствии прямого сигнала, при среднеквадратическом значении расширения задержки 250 нс.

Параметры лучевых компонент указаны в табл. 8.14. Все лучевые компоненты имеют классический доплеровский спектр за исключением первого луча модели *D* с пиком 10 дБ.

Таблица 8.14. Модели радиоканала внутри помещений

Номер луча	Модель <i>A</i>		Модель <i>B</i>		Модель <i>C</i>		Модель <i>D</i>		Модель <i>E</i>	
	Задержка, нс	Мощность, дБ	Задержка, нс	Мощность, дБ	Задержка, нс	Мощность, дБ	Задержка, нс	Мощность, дБ	Задержка, нс	Мощность, дБ
1	0	0	0	-2,6	0	-3,3	0	0	0	-4,9
2	10	-0,9	10	-3,0	10	-3,6	10	-10,0	10	-5,1
3	20	-1,7	20	-3,5	20	-3,9	20	-10,3	20	-5,2
4	30	-2,6	30	-3,9	30	-4,2	30	-10,6	40	-0,8
5	40	-3,5	50	0	50	0	50	-6,4	70	-1,3
6	50	-4,3	80	-1,3	80	-0,9	80	-7,2	100	-1,9
7	60	-5,2	110	-2,6	110	-1,7	110	-8,1	140	-0,3
8	70	-6,1	140	-3,9	140	-2,6	140	-9,0	190	-1,2
9	80	-6,9	180	-3,4	180	-1,5	180	-7,9	240	-2,1
10	90	-7,8	230	-5,6	230	-3,0	230	-9,4	320	0
11	110	-4,7	280	-7,7	280	-4,4	280	-10,8	430	-1,9
12	140	-7,3	330	-9,9	330	-5,9	330	-12,3	560	-2,8
13	170	-9,9	380	-12,1	400	-5,3	400	-11,7	710	-5,4
14	200	-12,5	430	-14,3	490	-7,9	490	-14,3	880	-7,3
15	240	-13,7	490	-15,4	600	-9,4	600	-15,8	1070	-10,6
16	290	-18,0	560	-18,4	730	-13,2	730	-19,6	1280	-13,4
17	340	-22,4	640	-20,7	880	-16,3	880	-22,7	1510	-17,4
18	390	-26,7	530	-24,6	1050	-21,2	1050	-27,6	1760	-20,9

Модели сверхширокополосных каналов. Все предшествующие модели используются для описания каналов со сравнительно узкими полосами частот радиоспектра. Сверхширокополосные сигналы могут занимать полосу частот от нуля до десятков ГГц. Спектральная плотность мощности этих сигналов очень мала, так что системы с указанными сигналами практически не создают помех узкополосным системам. Традиционные модели, представленные выше, не могут обеспечить адекватное описание каналов таких систем. Например, число лучевых компонент, имеющих одинаковое время задержки теперь невелико, так что центральная предельная теорема для их суммы теперь неприменима, в результате чего распределение Релея для огибающей суммы этих компонент использовать нельзя. Кроме того, импульсный отклик данного канала не является узкополосным, так как сигналы не имеют несущего гармонического колебания. Для таких каналов требуются новые модели. Стандарт IEEE 802.15.3a представляет один из возможных способов описания сверхширокополосных каналов для организации связи внутри помещений [8.7].

Модели исследовательских групп 3GPP и 3GPP-2. В предшествующих разделах рассматривались модели радиоканалов, имеющих один вход и один выход. Однако антенные решетки становятся неотъемлемыми элементами современных радиосистем цифровой связи. Для таких систем каналы имеют много входов и много выходов. Следовательно, для каналов требуются новые модели, которые естественно более сложные, чем рассмотренные выше. Работы по созданию таких моделей для сотовых систем третьего поколения в настоящее время ведутся в рамках двух проектов 3-rd Generation Partnership Project (3GPP). Список наиболее важных публикаций об этих работах можно найти в [8.7].

Глава 9

МЕТОДЫ ОБЕСПЕЧЕНИЯ ЭФФЕКТИВНОЙ РАБОТЫ СИСТЕМ РАДИОСВЯЗИ ПО МНОГОЛУЧЕВЫМ КАНАЛАМ

9.1. Классификация искажений сигналов, вызванных многолучевостью, и методов борьбы с ними

9.1.1. Классификация искажений

Условия многолучевого распространения радиоволн определяют частотную характеристику канала связи (ЧХК), причем если эти условия изменяются, то изменяется и частотная характеристика (см. гл. 8). В общем случае это явление приводит к искажениям сигналов, транслируемых по каналу связи. Искажения радиосигналов, вызванные многолучевым распространением радиоволн, зависят как от свойств канала связи, так и от характеристик используемых сигналов. Свойства канала определяются, как правило, моделями, описанными в гл. 8 и, в первую очередь, их числовыми характеристиками, такими как полоса когерентности по частоте Δf_k , время когерентности T_k , а также интервалом рассеяния во времени (расширением задержки) τ_s . Отметим, что из трех перечисленных параметров канала Δf_k и τ_s зависят, прежде всего, от структуры трассы распространения радиоволн в части размещения отражающих объектов, а T_k характеризуется движением приемной и передающей радиостанций или отражающих объектов трассы. Поэтому, если радиостанции и отражающие объекты неподвижны, так что трассы распространения всех лучей остаются неизменными, то и частотная характеристика канала неизменна. Это не исключает, однако, того факта, что под действием многолучевого распространения ЧХК в той или иной степени искажается. В частности, ее гарантированная полоса пропускания не превосходит величины Δf_k , а длительность импульсного отклика имеет порядок τ_s .

Если теперь радиостанции или отражатели перемещаются, то ЧХК изменяется во времени, так что приходится говорить уже о текущей частотной характеристике. Изменения параметров ЧХК приводят к дополнительным искажениям сигнала. Их именуют *замираниями*, поскольку часто они проявляются в существенных изменениях коэффициента передачи канала в процессе передачи. Время T_k , в течение которого коэффициент передачи успевает существенно измениться, может быть оценено некоторым временным интервалом. Из-за движения подвижного объекта за этот интервал расстояние между передатчиком и приемником радиолинии успевает измениться на величину порядка длины волны радиосигнала λ . При такой ситуации разности хода отдельных лучей также меняются примерно на величину λ . Это соответствует изменению фазовых сдвигов между лучами на величину, соизмеримую с 2π , что существенно изменит результат их суммирования. Таким образом, по порядку величины

$$T_k = \frac{\lambda}{v} = \frac{\lambda c}{vc} = 1 / \left(\frac{v}{c} f_0 \right) = \frac{1}{F_d}, \quad (9.1)$$

где v — радиальная скорость движения подвижного объекта; c — скорость света; f_0 — средняя частота радиосигнала; F_d — доплеровское смещение частоты радиолинии; это соотношение соответствует (8.39), (8.40).

На искажения сигналов цифровой манипуляции за счет многолучевости влияют, прежде всего, такие характеристики сигналов, как скорость изменения манипулируемых параметров, т.е. частота манипуляции B и, соответственно, ширина основного спектра сигнала Δf_c . В зависимости от соотношения параметров канала и сигнала различают следующие виды искажений сигнала:

- статические искажения, имеющие место при неизменных параметрах трассы распространения;
- дружные замирания, когда для всех частотных составляющих сигнала текущий коэффициент передачи канала оказывается приблизительно одинаковым; очевидно, что такое положение имеет место при

$$\Delta f_c \ll \Delta f_k; \quad (9.2)$$

- селективные замирания, когда отдельные составляющие спектра сигнала транслируются каналом с существенно различными коэффициентами передачи; они имеют место при

$$\Delta f_c > \Delta f_k; \quad (9.3)$$

- медленные замирания, когда длительность элементарного сигнала

$$T_3 = \frac{1}{B} \ll T_k, \quad (9.4)$$

так что за время целого ряда элементарных посылок ЧХК меняется незначительно;

- быстрые замирания, когда, наоборот,

$$T_3 \gg T_k. \quad (9.5)$$

Следует отметить, что понятия дружных и селективных замираний, так же как и медленных и быстрых, являются противоположными; естественно, что существовать одновременно замирания образующие эти пары не могут. В то же время дружные замирания могут быть как медленными, так и быстрыми. Это же утверждение относится и к селективным замираниям. Что касается статических искажений, то они по отношению к сигналу могут быть как дружными, так и избирательными. В первом случае при

$$\tau_3 \ll T_3 \quad (9.6)$$

они проявляются только в том, что коэффициент передачи канала принимает хотя и постоянное, но случайное значение. Во втором (τ_3 соизмеримо с T_3), кроме того, возникают еще и искажения формы элементарных посылок: при трансляции по каналу их форма изменяется в соответствии с ЧХК. В частности, длительность сигналов за счет переходных процессов, определяемых ЧХК, увеличивается, так что даже если на входе канала последовательно передаваемые элементарные сигналы следуют друг за другом без перекрытия во времени, то на выходе канала такое перекрытие возникает. Это явление, имеемое *межсимвольной интерференцией* (МСИ), приводит к тому, что полезный сигнал, принимаемый приемником на некотором тактовом интервале оказывается зависящим не только от ЭС, передававшегося на данном интервале, но и от ЭС, передававшихся ранее,

или, может быть, и позднее. Это обстоятельство должно учитываться при построении демодулятора приемника радиолинии. Действительно, как отмечалось в 3.1, оптимальность идеального приемника имеет место лишь если сигналы, передаваемые на различных тактовых интервалах, независимы. Там же указано, что, если такая зависимость фактически имеет место, необходимо переходить к приему «укрупненных» ЭС.

Классификация искажений сигналов в многолучевых каналах и условий их возникновения, иллюстрирует рис. 9.1.

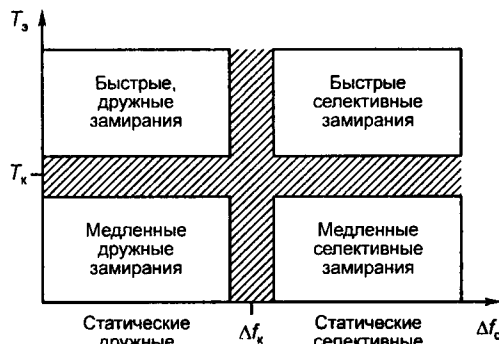


Рис. 9.1. Диаграмма искажений, вызванных многолучевостью

9.1.2. Классификация методов приема сигналов в многолучевых каналах

Помехоустойчивость приема сигналов при многолучевом канале всегда оказывается существенно меньше, чем для канала с постоянными параметрами, или, говоря иначе, при связи прямым лучом. Стремясь сократить соответствующие потери энергии радиолиний, используют различные специальные способы передачи и приема сигналов. Рациональный выбор этих способов для каждой конкретной радиолинии зависит от вида искажений, возникающих в ней за счет многолучевости, из числа тех, которые перечислены выше.

При создании любой радиолинии должны быть приняты все возможные меры для ослабления многолучевого распространения радиоволн на соответствующей трассе. Очевидным способом достижения этой цели является повышение направленности, прежде всего, приемной антенны радиостанции. Последнее сокращает количество отраженных лучей, принимаемых антенной и тем самым ослабляет многолучевость. Однако в ряде случаев, например, в системах подвижной службы, возможности следовать в данном направлении оказываются весьма ограниченными, так что реально в соответствующих каналах связи имеет место существенная многолучевость.

В ряде случаев удается кардинально решить проблему многолучевости путем разделения лучей на входе приемника. Такая возможность возникает при использовании в радиолинии широкополосных методов манипуляции, если параметры ШПС выбраны рационально в соответствии с характеристиками канала. В типовом случае использования ШПС (см. 6.2 и гл. 7) такой рациональный выбор определяется требованием

$$\tau_0 \ll \tau_s, \quad (9.7)$$

где τ_0 — длительность элементарного символа ПСП, использованной при формировании ШПС. При выполнении этого требования, подавляющая часть лучей всегда оказывается сдвинутой по отношению к прямому лучу на величину превосходящую τ_0 и потому оказываются ему почти ортогональными. Соответственно, на выходе устройства, осуществляющего свертку по частоте прямого луча, другие лучи не будут создавать заметного напряжения, так что и искажений, вызванных многолучевостью, не будет. Более того, если другие лучи имеют дискретные значения задержек, интервалы между которыми превосходят τ_0 , то каждый из них может быть выделен независимо от остальных. Далее все лучи существ-

венного уровня могут быть когерентно просуммированы, так что отраженные лучи не только не будут вредить приему прямого, но даже могут улучшать такой прием. Использование ШПС с разделением лучей служит основой технологии CDMA, предусмотренной стандартом IS-95 для соответствующих сотовых систем связи. (Подробнее см. в 11.2.)

К сожалению, применение ШПС с разделением лучей оказывается возможным далеко не всегда. Часто в распоряжении системы связи просто нет полосы частот, необходимой для сигналов ШПС с достаточно большой базой, определяемой требованием неравенства (9.7). Встречаются и иные причины, исключающие возможность реализации ШПС. В этих случаях разделение лучей оказывается невозможным, так что приходится обеспечивать прием искаженных многолучевостью сигналов.

Борьба с медленными дружными замираниями. Начнем с рассмотрения случая медленных дружных замираний. Основной используемый метод борьбы с замираниями сводится к разнесенному приему (передаче) сообщений. Общая идея метода сводится к организации вместо одной радиолинии, нескольких, работающих параллельно, по каждой из которых передаются сведения о транслируемом сообщении (одном и том же для всех этих радиолиний). В простейшем случае сообщение может просто повторяться по всем радиолиниям; в более сложном варианте, по каждой радиолинии передается свой набор сведений о сообщении, такой, что сообщение можно восстановить, если даже определенная часть радиолиний перестанет нормально функционировать, работая с повышенными вероятностями ошибок. Если сигналы во всех радиолиниях будут замирать синхронно, то никакой пользы описанная параллельная передача сообщения не принесет: часть сообщения, транслируемая в моменты, когда модуль коэффициента передачи канала опускается ниже некоторого минимально необходимого уровня будет потеряна во всех каналах и ее нельзя будет восстановить.

Другое дело, если замирания в параллельно работающих каналах будут независимы. Тогда с достаточно высокой вероятностью в любой момент времени найдутся каналы, по которым сообщение проходит достаточно надежно. Этими «хорошими» каналами в разные моменты времени окажутся разные каналы, но при рациональной обработке выходных потоков всех каналов можно будет восстановить переданное сообщение с достаточной надежностью на протяжении всего времени его передачи. Таким образом, для того, чтобы разнесение было эффективным, необходимо так формировать указанные параллельно работающие радиолинии, чтобы замирания в них были бы близкими к независимым. Набор указанных радиолиний может формироваться различными способами. Так, эти радиолинии могут создаваться путем разделения по частоте (см. 6.2) или на общей несущей методами пакетной передачи каналов с разделением их по времени (см. 6.3). Перечисленные способы разнесения именуется *разнесением по частоте и по времени* соответственно. Для создания таких радиолиний может использоваться и пространственное разделение (разнесение по пространству) или по поляризации. Теории и технике разнесенного приема (так именуется описанный метод, независимо от того, какими способами создаются параллельно действующие каналы) посвящен следующий раздел настоящей главы.

Борьба с селективными замираниями. Остановимся на статических искажениях сигналов, проявляющихся в виде МСИ. Как отмечалось выше, здесь необходимо использовать идеальный прием не отдельных ЭС, а «укрупненных» таких сигналов — цепочек ЭС. Непосредственная реализация такого приемника усложняется в степенной зависимости по мере роста числа ЭС в цепочке, становясь практически не реализуемой уже при значениях этой величины порядка 5 и более. Здесь приходит на помощь предложенный Витерби алгоритм, позволяющий осуществить идеальный прием сигналов с МСИ гораздо более простыми техническими средствами. Хотя с разновидностью алгоритма Витерби читатель уже встречался в гл. 4, где он использовался для декодирования сверточных

кодов, ниже в 9.3 этот алгоритм будет описан, но уже применительно к задаче идеального приема сигналов с МСИ.

Что касается селективных замираний, то они могут полностью исказить вид полезного ЭС, так что прием его станет невозможным даже и при полном отсутствии флуктуационных шумов. Поэтому при искажениях такого типа всегда необходимы специальные способы передачи и приема. Для борьбы с селективными искажениями применяются методы двух типов. Первый можно назвать методом устранения селективных замираний, а второй — методом их компенсации. Идея устранения селективных замираний сводится к следующему. Селективные замирания возникают в связи с чрезмерно широкой полосой манипулированного сигнала Δf_c , что определяется сравнительно высокой скоростью манипуляции этого сигнала B . Если теперь разделить весь поток символов сообщения на N подпотоков, то, используя тот же метод манипуляции, можно каждый подпоток передать на своей несущей при скорости манипуляции $B_n = B/N$ и полосе частот, занимаемой передачей каждого подпотока $\Delta f_{cn} = \Delta f_c/N$.

Таким образом, в идеале вместо передачи общего потока символов сообщения на общей несущей тот же поток в той же суммарной полосе частот будет передан с помощью N манипулированных сигналов, каждый на своей несущей. Последние будем именовать *парциальными сигналами*. Если N достаточно велико, так что $\Delta f_{cn} \ll \Delta f_c$, то в соответствии с (9.2) замирания для каждого парциального сигнала окажутся уже дружными, в результате специфические проблемы, связанные с селективными замираниями, отпадут. Это, однако, не означает, что влияющие замирания вообще исчезнет; в каждый момент времени часть парциальных сигналов из-за замираний не будет проходить, так что транслируемые ими подпотоки информации будут приниматься с недопустимо большими ошибками. Для того, чтобы такие ошибки могли быть исправлены, весь исходный информационный поток должен быть закодирован подходящим, достаточно эффективным помехоустойчивым кодом, а разделение на подпотоки должно производиться уже после кодирования и таким образом, чтобы близкие по временным положениям символы кода транслировались бы далеко отстоящими по частоте несущими. Тогда пропадание группы рядом стоящих по частоте парциальных сигналов, вызванное замиранием, не приведет к появлению пачек ошибок в кодовой комбинации и сможет быть эффективно исправлено кодом.

Представленная выше идея устранения селективных замираний является достаточно очевидной и тот факт, что до недавнего времени она не использовалась на практике связан с тем, что его реализация традиционными схемотехническими приемами оказывается слишком громоздкой. Легко представить, какой сложности аппаратуру пришлось бы создавать скажем, при $N = 1000$. Достаточно отметить, что соответствующий модем должен был бы иметь синтезатор, генерирующий одновременно 1000 несущих частот, 1000 модуляторов и столько же демодуляторов, не говоря уже о большом числе других устройств и блоков. Однако современная цифровая технология, основанная на быстром преобразовании Фурье, позволяет весьма компактно решать эту задачу, используя строго когерентные парциальные сигналы и специальные приемы для исключения влияния МСИ. В результате удается достичь весьма высоких показателей как по спектральной, так и по энергетической эффективности. Соответствующий метод манипуляции получил название Orthogonal Frequency Division Multiplexing (OFDM) (подробнее см. в 9.5).

Если описанный метод устранения селективных замираний основан на специальном выборе методов передачи и приема сигналов, то метод их компенсации использует только специальное усовершенствование метода приема при обычных методах передачи (манипуляции). Идея метода компенсации сводится к измерению текущей ЧХК и ее соответствующей коррекции. Текущая ЧХК измеряется периодически с интервалом, в течение которого эта

характеристика не успевает еще слишком существенно измениться. С этой целью в транслируемую по радиолинии последовательность бит периодически вставляется маркерная комбинация, вид которой заранее известен на приемной стороне. Такую комбинацию называют еще контрольным, или обучающим словом. При демодуляции вначале выделяется сигнал на временном интервале, несущем эту комбинацию, и его форма сопоставляется с той, которую он должен был бы иметь при отсутствии замираний (поскольку содержание обучающего слова заранее известно, отыскание последнего не составляет труда). Далее определяется текущая ЧХК, которую мы обозначим через $K_j(i\omega)$, где i — мнимая единица; j — номер измерения, в результате которого получена данная характеристика.

Использование этих данных при демодуляции может вестись двумя путями. Первый сводится к коррекции текущей ЧХК путем формирования специальной корректирующей цепи, через которую должен пропускаться входной сигнал приемника, прежде чем он поступит на демодулятор. В полосе частот, занимаемой полезным сигналом, частотная характеристика корректирующей цепи $K_K(i\omega)$ должна быть такой, чтобы приближенно выполнялось соотношение $K_j(i\omega)K_K(i\omega) = \text{const}$.

Во втором варианте аналогично формируется требуемая корректирующая цепь, но принимаемый сигнал через нее не пропускается, а коррекции подвергаются лишь образцы укрупненных ЭС (цепочек ЭС), используемых при демодуляции в идеальном приемнике. На первый взгляд может показаться, что между обоими вариантами нет принципиальных различий. Что это не совсем так, становится ясно, если учесть, что входной сигнал приемника неотделим от шумов. При пропускании данного сигнала через корректирующую цепь через нее проходят и шумы. И если эта цепь существенно усиливает некоторые участки спектра сигнала, которые оказались подавленными в канале связи вследствие замираний, то одновременно она резко повышает шумовой вклад указанных участков спектра, так что такого рода коррекция может оказаться даже вредной. Данная опасность отсутствует, если коррекции подвергаются образцы укрупненных ЭС. Однако во втором варианте демодулятор должен строиться так, чтобы учитывать МСИ, имеющие место в принимаемом сигнале, в то время как в первом корректирующая цепь в идеале устраняет и МСИ, так что демодулятор резко упрощается.

Отметим, что в обоих случаях совершенно не обязательно физически формировать корректирующие цепи; их функции могут выполняться посредством соответствующего пересчета координат цифровых представлений соответствующих сигналов. Устройство приемника, выполняющее тем или иным путем функции определения текущих ЧХК, и, как следствие, требуемых характеристик корректирующих цепей, а также и процедуры коррекции форм принимаемого сигнала или образцов укрупненных ЭС, именуется *эквалайзером*. Минимальный объем сведений об этом весьма сложном устройстве представлен в 9.4.

9.2. Медленные дружные замирания и разнесенный прием

9.2.1. Предпосылки и исходные соотношения

Как видно из материала предыдущей главы, статистическое описание замираний в каналах связи достаточно сложно, причем оно изменяется в весьма широких пределах для различных, встречающихся на практике каналов. Тем более сложную, а часто и невыполнимую задачу представляет достаточно точная аналитическая оценка эффективности тех или иных методов борьбы с замираниями. Приближающиеся к реальным, модели каналов используются в настоящее время при моделировании каналов на ЭВМ и испытаниях оборудования; при аналитических же расчетах опираются, как правило, на упрощенные модели, при выборе кото-

рых приходится в качестве одного из основных факторов учитывать удобство такого их использования. Естественно, что и упрощенные модели должны, пусть в грубой форме, но отражать реальные характеристики каналов.

Отметим, что даже и при далеко идущих упрощениях моделей получить компактные и удобные для использования аналитические результаты по рассматриваемым ниже проблемам удастся далеко не всегда. Поэтому ниже при оценке эффективности тех или иных методов борьбы с замираниями широко применяются численные расчеты. Приведем вначале некоторые исходные соотношения, которые будут далее широко использоваться в настоящем разделе.

Изменения квадратурных компонент сигнала на входе приемника вследствие замираний можно описать, считая их пропорциональными случайным величинам μ_I и μ_Q соответственно, представляющими собой коэффициенты передачи канала для квадратурных компонент сигнала. Математические ожидания этих величин определяют уровень регулярной компоненты сигнала на входе приемника

$$M_p = \sqrt{(\overline{\mu_I})^2 + (\overline{\mu_Q})^2}. \tag{9.8}$$

Остальную часть принимаемого сигнала образуют диффузные его компоненты. Суммарная мощность всех компонент сигнала на входе приемника пропорциональна величине

$$\mu^2 = \mu_I^2 + \mu_Q^2, \tag{9.9}$$

Далее

$$M_\phi^2 = \overline{\mu^2} = D\mu_I + (\overline{\mu_I})^2 + D\mu_Q + (\overline{\mu_Q})^2 = M_p^2 + M_\phi^2, \tag{9.10}$$

где $D\xi$ — дисперсия случайной величины ξ ; $M_\phi^2 = D\mu_I + D\mu_Q$ — суммарная мощность всех диффузных компонент сигнала на входе приемника.

В качестве грубой, но удобной для расчетов аппроксимации плотности распределения вероятностей случайной величины μ ниже принято распределение Райса (см. (1.43)). С учетом обозначений настоящего пункта, оно принимает вид

$$p_\mu(x) = \frac{2x}{M_\phi^2} \exp\left(-\frac{x^2 + M_p^2}{M_\phi^2}\right) I_0\left(2\frac{xM_p}{M_\phi^2}\right) \text{ при } x \geq 0 \text{ и } p_\mu(x) = 0 \text{ при } x < 0. \tag{9.11}$$

Величина μ призвана отражать лишь относительные изменения коэффициента передачи канала под действием замираний, а не абсолютные его значения. Поэтому удобно ее масштаб выбрать таким, чтобы всегда $M_0 = 1$. Тогда в соответствии с (9.10) получим

$$M_\phi^2 = \frac{1}{k^2 + 1} \text{ и } M_p^2 = \frac{k^2}{1 + k^2}, \tag{9.12}$$

где $k = M_p/M_\phi$ — отношение эффективных напряжений регулярной и диффузной составляющих сигнала на входе приемника. С учетом такой нормировки вместо (9.11) получим

$$p_\mu(x) = 2xk_1^2 \exp(-x^2k_1^2 - k^2) I_0(2xkk_1), \tag{9.11a}$$

где $k_1^2 = 1 + k^2$. Распределение для различных значений k представлено на рис. 9.2.

Отметим, что, как легко показать, асимптотически, т.е. при $k \rightarrow \infty$, распределение (9.11a) сходится к нормальному с математическим ожиданием $M_p = k/k_1$ и дисперсией $\sigma^2 = M_\phi / 2 = 1/(2k_1^2)$.

При замираниях в соответствии с изменением величины μ происходит и изменение мощности полезного сигнала на входе приемника, а следовательно, и нормированного отношения сигнал/шум h_ϕ (см. (3.18)). Последнее может быть записано в виде $h_\phi^2 = \mu^2 h_{\phi 0}^2$, где

h_{60}^2 — представляет собой нормированное отношение сигнал/шум на входе приемника, рассчитанное в предположении, что мощность полезного сигнала на этом входе равна ее среднему значению при замираниях. Таким образом, h_{60}^2 — некоторая константа, в отличие от h_6 , являющейся случайной величиной.

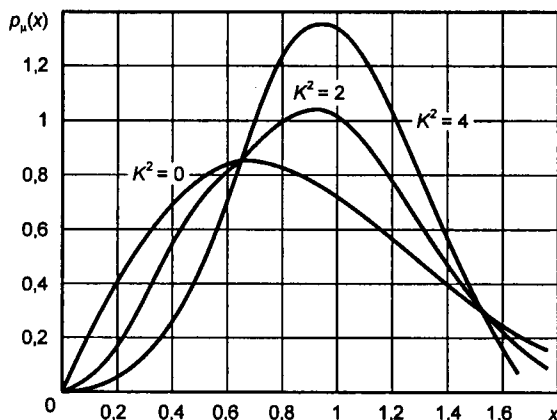


Рис. 9.2. Плотность распределения вероятностей μ для райсовских замираний

Поскольку при замираниях нормированное отношение сигнал/шум на входе приемника радиолинии случайно, то случайной оказывается и вероятность ошибки в сообщении на выходе демодулятора. Ниже, в настоящем разделе ограничимся примерами, когда сообщение передается в двоичной форме методом ОФМ. При этом будем полагать, как правило, что в качестве демодулятора используется некогерентный идеальный приемник (см. 3.3). Такой выбор определяется только стремлением упростить вычисления, в то время как некоторые потери, связанные с заменой когерентного приема некогерентным не идут, как мы увидим ниже, ни в какое сравнение с потерями от замираний, для выявления которых и будут проводиться нижеследующие расчеты.

В соответствии с (3.40) и приведенными выше обозначениями вероятность ошибки на бит как случайная величина определяется соотношением

$$q_{\text{ошб}} = 0,5 \exp - h_6^2 = 0,5 \exp - \mu^2 h_{60}^2. \quad (9.13)$$

Среднее значение этой вероятности может быть получено путем ее усреднения по распределению (9.11а). Используя соответствующие табличные интегралы, такое усреднение удастся выполнить, и оно приводит к следующему соотношению, определяющему среднюю вероятность ошибки на выходе демодулятора:

$$P_{\text{ошб}} = 0,5 \frac{k_1^2}{k_1^2 + h_{60}^2} \exp - \frac{k^2 h_{60}^2}{k_1^2 + h_{60}^2}. \quad (9.14)$$

Ниже нам потребуется и некоторое очевидное обобщение данного соотношения. Именно, если некая функция от μ имеет вид $y = 0,5a \exp - b\mu h_{60}^2$, где a и b — некоторые неотрицательные константы, то среднее значение y по распределению (9.11а) составит

$$\bar{y} = 0,5 \frac{ak_1^2}{k_1^2 + bh_{60}^2} \exp - \frac{k^2 bh_{60}^2}{k_1^2 + bh_{60}^2}. \quad (9.15)$$

9.2.2. Помехоустойчивость при отсутствии разнесения

Если в канале с медленными замираниями осуществляется посимвольная передача сообщений без помехоустойчивого кодирования, то средняя вероятность ошибки на бит сообщения $P_{\text{ошб}}$ определяется соотношением (9.14). Рассмотрим некоторые частные случаи.

Пусть регулярная компонента сигнала на входе приемника отсутствует, т.е. в канале связи имеет место чисто диффузное распространение. Такой канал именуют обычно релейским, в отличие от общего случая райсовского канала.

Для релейского канала $k = 0$, так что

$$P_{\text{ошб}} = 0,5 \frac{1}{1 + h_{60}^2}. \tag{9.16}$$

Соотношение (9.16) указывает на столь значительные потери в помехоустойчивости, вызванные замираниями, что их с полным правом можно назвать катастрофическими. Действительно, для обычно требуемой на практике вероятности ошибки 10^{-5} (9.16) дает $h_{60} = 47$ дБ, в то время как при отсутствии замираний в соответствии с (9.13) достаточно иметь $h_{60} = 10$ дБ. Таким образом, энергетический проигрыш, вызванный замираниями, составляет 37 дБ, или 5 тысяч раз! Ясно, что в релейском канале совершенно необходимо принимать все возможные меры для ослабления влияния замираний. Если имеет место райсовский канал, то по мере увеличения k потери, вызванные замираниями, сокращаются (см. рис. 9.3) и при $k \rightarrow \infty$ стремятся к 0 дБ. Для достаточно малых вероятностей ошибок и соответственно достаточно больших значений h_{60} , если $h_{60} \gg k$ из (9.14) получаем следующее асимптотическое соотношение

$$P_{\text{ошб}} = 0,5 \frac{k_1^2}{h_{60}^2} \exp - k^2. \tag{9.17}$$

Сопоставляя его с предыдущим соотношением приходим к выводу, что при типовых значениях k порядка (5–7) дБ потери в райсовском канале, хотя и значительно сокращаются по сравнению с релейским (на 7–14 дБ), но все-таки остаются еще весьма значительными. И только для k -го порядка 12 дБ и более потери энергии радиолинии, вызванные медленными замираниями, при посимвольном приеме опускаются до значений порядка нескольких дБ и менее. Помехоустойчивость посимвольной передачи сообщений при медленных замираниях, представлена на рис. 9.3, кривые которого соответствуют соотношению (9.14).

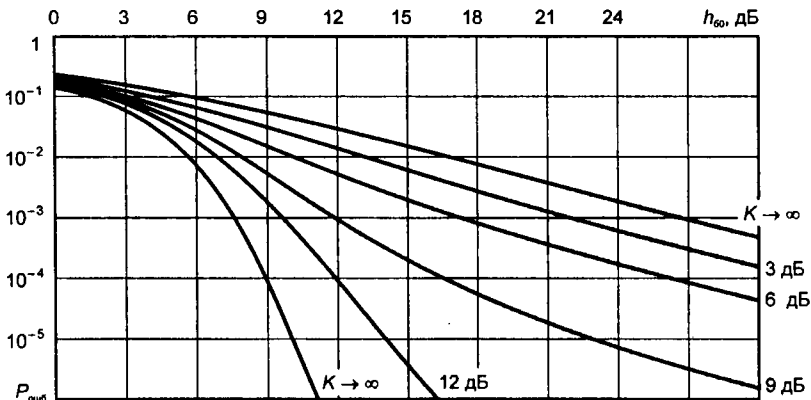


Рис. 9.3. Помехоустойчивость посимвольной передачи сообщений в канале с замираниями

Для повышения помехоустойчивости передачи сообщений в канале с замираниями можно попытаться применить тот же способ, который является весьма эффективным в канале при отсутствии замираний (при связи прямым лучом). Речь идет об использовании помехоустойчивого кодирования, которому посвящена гл. 4. Из описанных там кодов для количественной оценки их эффективности в каналах с медленными замираниями выбрано два типовых. Это один из самых эффективных коротких блочных кодов — код Голея (23,12,7) и сверточный код с глубиной кодового ограничения $K = 7$ и скоростью $R = 1/2$. Принимается, что код Голея декодируется посимвольно, а сверточный код — в целом по алгоритму Витерби.

Будем полагать, что помехоустойчивое кодирование используется в сочетании с ФМ2 или ФМ4 при когерентном приеме. Поскольку замирания в канале медленные, нормированное отношение сигнал/шум h_6 , хотя и является случайной величиной, но сохраняется почти неизменным на протяжении времени, в течение которого передается достаточно большое число бит. Если при этом символы кода передаются в их естественном порядке (по мере формирования), то можно считать, что все кодовые биты, участвующие в текущих процедурах декодирования, принимаются при одном и том же отношении h_6 . Это позволяет использовать следующую методику оценки вероятности ошибки при передаче кодированных сообщений по каналам с медленными замираниями: за основу принимается зависимость вероятности ошибки на бит сообщения от отношения h_6 в канале с постоянными параметрами при рассматриваемом коде и способе декодирования. Запишем эту зависимость в виде $q_{\text{ошб}}(h_6) = q_{\text{ошб}}(\mu h_{60})$. Таким образом, $q_{\text{ошб}}$ оказывается случайной величиной. Усредненную вероятность ошибки $P_{\text{ошб}} = \overline{q_{\text{ошб}}(\mu h_{60})}$ найдем, используя распределение (9.11а).

Выполним эти вычисления для кода Голея. В качестве исходной зависимости $q_{\text{ошб}}(h_6)$ используем соответствующую кривую на рис. 4.19. Эта кривая повторена на рис. 9.4, а.

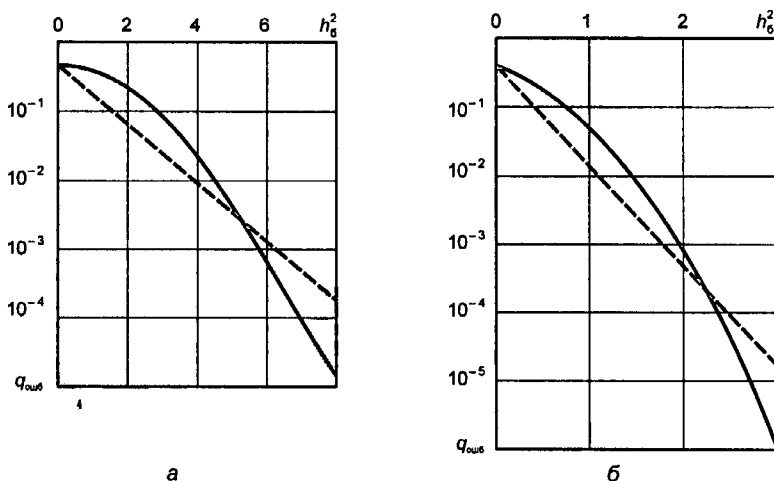


Рис. 9.4. Зависимость $q_{\text{ошб}}(h_6)$ для кода Голея (а) и сверточного кода (б) и их соответствующие аппроксимации

Нам остается усреднить значения $q_{\text{ошб}}$. При реальных расчетах это следовало бы делать численными методами. Однако здесь для получения более наглядных аналитических результатов поступим иначе. Грубо аппроксимируем зависимость $q_{\text{ошб}}(h_6)$, представленную сплошной линией на рис. 9.4, в виде штриховой линии, т.е. примем

$$q_{\text{ошб}} = 0,5 \exp -1,6 h_6^2 = 0,5 \exp -1,6 \mu h_{60}^2. \quad (9.20)$$

Усреднение данной величины по распределению (9.11а) с помощью соотношения (9.15) дает

$$P_{\text{ошб}} = 0,5 \frac{k_1^2}{k_1^2 + 1,6h_{60}^2} \exp - \frac{1,6k^2 h_{60}^2}{k_1^2 + 1,6h_{60}^2}. \quad (9.21)$$

Аналогично рассмотрим и случай сверточного кода. В качестве базовой примем кривую на рис. 4.28, относящуюся к глубине кодового ограничения $K = 7$ и скорости кода $R = 1/2$. На рис. 9.4, б представлена эта зависимость и ее аппроксимация (штриховая).

Аналитическая запись аппроксимации имеет вид

$$q_{\text{ошб}} = 0,5 \exp - 2,53h_6^2 = 0,5 \exp - 2,53\mu h_{60}^2. \quad (9.22)$$

Соответственно усредненная ошибка оказывается:

$$P_{\text{ошб}} = 0,5 \frac{k_1^2}{k_1^2 + 2,53h_{60}^2} \exp - \frac{2,53k^2 h_{60}^2}{k_1^2 + 2,53h_{60}^2}. \quad (9.23)$$

Сопоставление соотношений (9.14), (9.21) и (9.23) показывает, что помехоустойчивое кодирование, реализованное так как его используют в каналах с постоянными параметрами, в каналах с медленными замираниями, хотя и обеспечивает энергетический выигрыш несколько дБ, но ни в коей мере не может компенсировать возникающие потери помехоустойчивости. Это означает, что в каналах с замираниями должны реализовываться специфические методы для компенсации возникающих потерь.

9.2.3. Кодирование с разнесением во времени

Для каналов с медленными замираниями такую специфическую группу методов (см. 9.2.1) образует разнесенный прием. Одним из методов данной группы является помехоустойчивое кодирование с разнесением во времени, именуемое чаще *кодированием с перемежением*. Идея перемежения сводится к тому, что кодовые символы передаются в канал связи не в том порядке, в котором они возникают на выходе кодера, а после перестановки, причем такой, чтобы близлежащие по моментам возникновения символы, различались по времени передачи в канал не менее, чем на время когерентности канала T_K .

Технически такая перестановка кодовых символов может быть выполнена с помощью матрицы памяти, двоичные элементы памяти которой набраны в N строк при M столбцах. Символы с выхода кодера записываются, скажем, в строки этой матрицы последовательно в одну за другой. Считывание их в канал связи осуществляется только после заполнения всей матрицы и по столбцам также последовательно один за другим. Возможны варианты, когда используемые при записи строки и столбцы при считывании чередуются в том или ином установленном порядке. На приемной стороне принятые кодовые символы записываются в аналогичную матрицу памяти на те же позиции, которые они занимали в передающей матрице. После заполнения приемной матрицы кодовые символы подаются на вход декодера в той последовательности, которую они имели на выходе кодера. Из изложенного принципа ясно, что максимальная задержка в доставке информации, вызванная перемежением, равна суммарной длительности $2NM$ кодовых символов

Попытаемся теперь оценить помехоустойчивость передачи информации в канале с медленными замираниями при использовании кодов с перемежением.

В общем случае демодулятор приемника радиолинии выдает данные о коэффициенте правдоподобия для очередного принимаемого кодового символа. При цифровой реализации демодулятора эти данные содержат знак этого коэффициента и значение его модуля, кото-

рое определяется двоичным числом заранее выбранной разрядности n . Пусть для определенности знаку плюс соответствует прием символа «1», а минус — символа «0». Можно вместо знака добавить в числовое значение коэффициента правдоподобия еще один разряд (старший), полагая его равным единице при знаке плюс и нуль при знаке минус. Теперь старший разряд коэффициента правдоподобия непосредственно совпадает с жестким решением демодулятора. Если декодер использует жесткие решения, то при построении демодулятора можно принять $n = 0$, ограничиваясь выдачей только этого старшего разряда. Если имеет место мягкое декодирование, то достаточно (см. гл. 4) принять $n = 3 - 4$.

Условимся в общем случае считать ошибкой на выходе демодулятора ошибочное воспроизведение старшего разряда коэффициента правдоподобия, т.е. когда при передаче кодового символа «1» этот разряд принимает значение нуль, и наоборот. Вероятность такой ошибки в каналах с медленными замираниями $q_{\text{ошд}}$ изменяется в соответствии с изменением величины нормированного отношения сигнал/шум. Нормировку здесь удобно проводить относительно символов кода. Полученное в результате отношение $h_{\text{бк}}^2 = Rh_{\text{б0}}^2$, где R — скорость кода (см. 4.2) и определяет вероятность $q_{\text{ошд}}$. Изменение $q_{\text{ошд}}$ во времени должно учитываться в случае, когда перемежение при кодировании не используется, ибо вся кодовая комбинация или эквивалентный ей отрезок кодированной последовательности принимается при такой вероятности ошибки. Именно так и было сделано выше при анализе этого случая. Если же используется достаточно глубокое перемежение, то вероятности ошибки на выходе демодулятора для всех символов кода оказываются независимыми и постоянными и могут быть получены путем усреднения $q_{\text{ошд}}$:

$$P_{\text{ошд}} = q_{\text{ошд}}(h_{\text{бк}}) = q_{\text{ошб}}(\mu\sqrt{Rh_{\text{б0}}}). \quad (9.24)$$

В соответствии с (9.15) найдем при перемежении

$$P_{\text{ошд}} = 0,5 \frac{k_1^2}{k_1^2 + Rh_{\text{б0}}^2} \exp - \frac{k^2 Rh_{\text{б0}}^2}{k_1^2 + Rh_{\text{б0}}^2}. \quad (9.25)$$

С другой стороны, для каждого конкретного кода и способа его декодирования можно установить связь между вероятностями ошибок на выходе демодулятора и декодера. Для кода Голея ошибки на выходе декодера происходят с вероятностью порядка 0,5 только тогда, когда неправильно принята кодовая комбинация. Последнее происходит тогда и только тогда, когда, за время длительности такой комбинации демодулятор выдает четыре ошибки или более. Следовательно, вероятность ошибки на выходе декодера приближенно может быть выражена через вероятность ошибки на его входе соотношением

$$P_{\text{ошб}} = 0,5 \sum_{j=4}^{23} C_{23}^j P_{\text{ошд}}^j (1 - P_{\text{ошд}})^{23-j}. \quad (9.26)$$

Эта зависимость представлена табл. 9.1.

Таблица 9.1. Связь вероятностей ошибок на входе и выходе декодера для кода Голея при жестком декодировании

$P_{\text{ошд}}$	10^{-1}	$8 \cdot 10^{-2}$	$6 \cdot 10^{-2}$	$4 \cdot 10^{-2}$	$3 \cdot 10^{-2}$	$2 \cdot 10^{-2}$	$1,5 \cdot 10^{-2}$	10^{-2}	$6 \cdot 10^{-3}$
$P_{\text{ошб}}$	$2,8 \cdot 10^{-2}$	$1,8 \cdot 10^{-2}$	$7 \cdot 10^{-3}$	$1,8 \cdot 10^{-3}$	$6 \cdot 10^{-4}$	$1,45 \cdot 10^{-4}$	$5 \cdot 10^{-5}$	$1,2 \cdot 10^{-5}$	$5 \cdot 10^{-6}$

Расчет помехоустойчивости при кодировании с перемежением с использованием приведенных соотношений может быть произведен следующим образом. Для заданных характеристик замираний канала (величине k) и среднем отношении сигнал/шум на входе приемника

h_{60}^2 по формуле (9.25) определяется вероятность ошибки на выходе демодулятора $P_{\text{ошб}}$. Далее по табл. 9.1 вычисляется вероятность ошибки на выходе декодера.

Расчеты для других кодов и других способов декодирования при достаточно глубоком перемежении отличаются только видом зависимости, отображаемой табл. 9.1. Так, для сверточного кода с $K = 7$ и $R = 1/2$ при мягком декодировании по алгоритму Витерби зависимость между вероятностями ошибок на входе и выходе декодера принимает вид, представленный в табл. 9.2.

Таблица 9.2. Связь вероятностей ошибок на входе и выходе декодера для сверточного кода при мягком декодировании

$P_{\text{ошд}}$	0,16	0,11	$8 \cdot 10^{-2}$	$6,7 \cdot 10^{-2}$	$5,6 \cdot 10^{-2}$	$4,8 \cdot 10^{-2}$	$4 \cdot 10^{-2}$
$P_{\text{ошб}}$	$8 \cdot 10^{-2}$	$1 \cdot 10^{-2}$	$1 \cdot 10^{-3}$	$1,5 \cdot 10^{-4}$	$7 \cdot 10^{-5}$	$7 \cdot 10^{-6}$	$1 \cdot 10^{-6}$

Табл. 9.2 построена по данным рис. 4.27 следующим образом. Используется кривая этого рисунка, соответствующая коду с $R = 1/2$. Задаемся $P_{\text{ошб}}$ и по этой кривой определяем соответствующее значение h_6 . Умножая его на \sqrt{R} , находим $h_{6к}$, а по нему в соответствии с данными о потенциальной помехоустойчивости используемого метода манипуляции (см. 3.2) находим значение $P_{\text{ошд}}$.

Помехоустойчивость передачи сообщений по каналу с медленными замираниями с использованием кода Голея при перемежении и жестком декодировании, определенная в соответствии с описанной выше методикой, представлена на рис. 9.5, а.

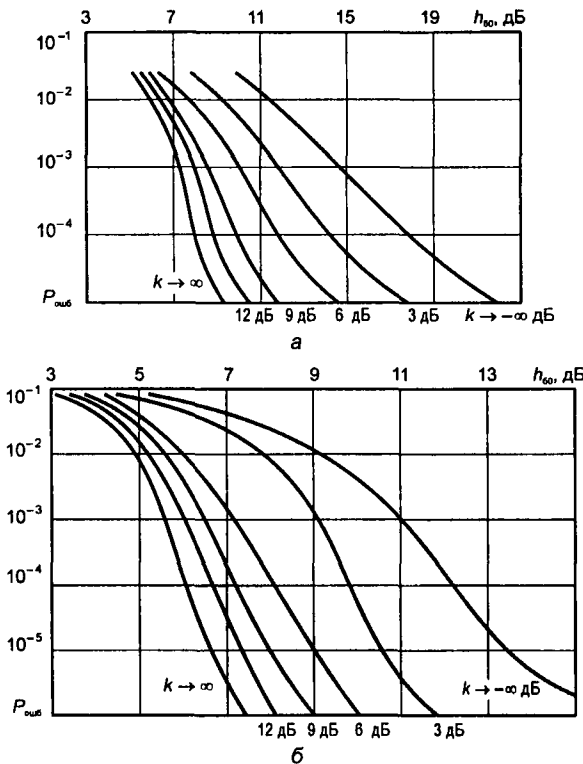


Рис. 9.5. Помехоустойчивость при коде Голея с перемежением (а) и сверточном коде с перемежением (б)

Сопоставление данных на рис. 9.5 с данными рис. 9.3 и формулы (9.16) показывает, что кодирование с перемежением обеспечивает в каналах с медленными замираниями весьма существенные энергетические выигрыши и при том тем большие, чем меньше k и требуемая вероятность $P_{\text{ошб}}$. Так, для релейского канала и $P_{\text{ошб}} = 10^{-5}$ выигрыш составляет 22 дБ, а для $P_{\text{ошб}} = 10^{-3}$ — 12,5 дБ. Для $k = 9$ дБ и $P_{\text{ошб}} = 10^{-5}$ выигрыш — 12 дБ, а для $P_{\text{ошб}} = 10^{-3}$ — всего 3,5 дБ. Во всех случаях он соизмерим с потерями, вызванными замираниями, и позволяет существенно сократить эти потери.

Представленная выше методика позволяет оценить помехоустойчивость и при сверточном коде с декодированием в целом по алгоритму Витерби. Соответствующие зависимости представлены на рис. 9.5, б.

Сопоставляя результаты с приведенными выше данными, относящимися к случаю, когда кодирование не используется, приходим к качественно тем же заключениям, что и для кода Голея. Количественные значения энергетических выигрышей здесь еще более впечатляющие: для релейского канала для $P_{\text{ошб}} = 10^{-5}$ выигрыш составляет 30,5 дБ, а для $P_{\text{ошб}} = 10^{-3}$ — 16 дБ; для $k = 9$ дБ и $P_{\text{ошб}} = 10^{-5}$ и $P_{\text{ошб}} = 10^{-3}$ он уменьшается до 16 дБ и 5,5 дБ.

По поводу данных рис. 9.5 сделаем следующее замечание: у внимательного читателя может вызвать недоумение тот факт что представленные кривые для $k \rightarrow \infty$, т.е. для канала без замираний, несколько отличаются от относящихся к этому случаю зависимостей, приведенных в гл. 4 (см. рис. 4.17 и 4.25, б). Данное обстоятельство связано с тем, что рассматривается вариант, когда используется ОФМ при некогерентной демодуляции, в то время как в гл. 4 изучалось использование когерентной демодуляции в сочетании с ФМ.

Итак, приведенные выше результаты позволяют сделать вывод о том, что кодирование с перемежением является эффективным средством снижения энергетических потерь, вызванных медленными дружными замираниями, поэтому, по возможности, должно всегда использоваться в таких каналах. Основной проблемой, возникающей при перемежении, оказывается дополнительная задержка в передаче информации, которая, как следует из соображений, приведенных в начале данного пункта, должна быть не менее $2T_k$. Так, для случая, например, телефонной связи эта задержка может оказаться уже существенной, поскольку дополнительное запаздывание порядка 100 мс уже ощущается при телефонном обмене как некая «заторможенность» абонента. С другой стороны, время когерентности канала T_k зависит, в первую очередь, от скорости движения организующих канал связи радиостанций, поэтому не может быть предсказано заранее. Для неподвижных радиостанций он вообще может оказаться неограниченно большим. Исходя из данного обстоятельства стремятся, когда это возможно, ограничить величину T_k . Это может быть достигнуто, в частности, путем регулярной смены номинала несущей частоты радиoliniии. Если сдвиг частоты превосходит интервал корреляции канала по частоте Δf_k , то значения μ до скачка частоты и после него будут статистически независимыми, так что время когерентности канала не будет превосходить временного интервала между скачками частоты. Таким образом, удастся ограничить сверху значение T_k , а следовательно, и глубину перемежения при кодировании *MM*.

Отметим, что использование скачков по частоте не означает, вообще говоря, снижения эффективности использования спектра рабочих частот. Эта процедура может сводиться просто к тому, что радиостанции системы связи, использующей некоторый интервал частот, должны с определенным тактом и по определенному закону обмениваться рабочими частотами, что не требует какого-либо расширения указанного интервала частот.

9.2.4. Другие варианты разнесенного приема

Как отмечалось выше, на практике применяют и другие методы разнесенного приема, использующие разнесение по частоте, пространству и т.д. Из этих методов наибольшее распространение, особенно в системах фиксированной службы, получил вариант с разнесением по пространству приемных антенн. В этом варианте передача сообщения осуществляется полностью аналогично случаю канала без замираний. Для приема, используется не одна, а несколько (n) приемных антенн, расположенных на расстояниях, превосходящих интервал корреляции канала по пространству. Замирания сигналов на выходах этих антенн должны оказаться взаимно независимыми.

Сигналы с выходов всех приемных антенн (мы будем говорить ниже, со всех ветвей разнесения) поступают на общее устройство их совместной обработки, которое и формирует демодулированное сообщение. Возможны различные алгоритмы совместной обработки, причем их выбор существенно влияет на эффективность борьбы с замираниями. Эта эффективность зависит, в частности, от объема сведений о сигнале каждой ветви, которые используются при обработке. К таким сведениям относятся: текущее отношение сигнал/шум ветви h_j ($j = 1, 2, \dots, n$), спектральная плотность N_{0j} шумов в ней, взаимный сдвиг фаз между полезными сигналами отдельных ветвей и др. Все эти параметры не бывают известны заранее и должны определяться в реальном времени, для чего в устройстве совместной обработки включают соответствующие анализаторы. Очевидно, что определение указанных параметров в реальных каналах связи представляет собой достаточно сложную задачу, сопряженную с необходимостью реализации весьма сложных устройств, так что, чем больший объем сведений о сигналах в ветвях разнесения используется при совместной обработке, тем более сложным оказывается соответствующее оборудование. Это и объясняет стремление к упрощению алгоритмов совместной обработки, чем и определяется рассмотрение в нижеследующем материале нескольких таких алгоритмов.

Начнем с анализа оптимального алгоритма совместной обработки, именуемого *когерентным сложением с весами*. Речь идет о том, что в устройстве совместной обработки напряжения ветвей $z_j(t)$ когерентно (по полезному сигналу) суммируются, причем каждому напряжению приписывается определенный вес β_j , так что суммарный сигнал

$$Z(t) = \sum_{j=1}^n \beta_j z_j(t). \quad (9.27)$$

Суммарное напряжение подается на обычный демодулятор для выделения сообщения. Поясним смысл введения весов β_j . Напряжения ветвей содержат как полезный сигнал $x_j(t)$, так и шумы $\xi_j(t)$. То же относится и к суммарному напряжению $Z(t)$. В каждой ветви в каждый данный момент времени имеет место свой, отображающий замирания, коэффициент передачи канала μ_j . В тех каналах, где в данный момент коэффициент передачи оказывается много меньше единицы, напряжение ветви содержит почти исключительно шум, так что добавление этого напряжения к (9.27) лишь снизило бы отношение сигнал/шум в суммарном напряжении. Если же для такого сигнала принимать вес достаточно малым, то такого ухудшения не произойдет.

Рассмотрим вопрос об оптимальном выборе весовых коэффициентов β_j . Обозначим текущее значение отношений сигнал/шум по мощности в ветвях через $h_j^2 = \mu_j^2 h_{\text{ср}}^2$, где $h_{\text{ср}}^2$ — среднее значение отношения мощностей сигнала и шума в ветви при $\mu^2 = \overline{\mu_j^2} = 1$. Поскольку ветви предполагаются подобными, принимается, что для всех них $h_{\text{ср}}^2$ одинаково. Что касается суммарного напряжения $Z(t)$, то обозначим, отношение в нем сигнал/шум по мощно-

сти через h^2 . Задача выбора весов β_j заключается в нахождении таких их значений, при которых h^2 оказывается максимальным. Отметим сразу что, без ограничения общности один из весов можно выбрать произвольно, вычисляя остальные относительно него. Такой выбор не повлияет на достижимое значение h^2 , а определит лишь результирующий уровень суммарного напряжения. Выберем $\beta_1 = 1$.

Примем мощности шумов во всех каналах одинаковыми и равными $P_{\text{ш}}$. Тогда мощность полезного сигнала в напряжении j -й ветви $P_{\text{с}j} = h_j^2 P_{\text{ш}}$, а

$$h^2 = \frac{\left(\sum_{j=1}^m \beta_j h_j \sqrt{P_{\text{ш}}} \right)^2}{\sum_{j=1}^n \beta_j^2 P_{\text{ш}}} = \frac{\left(\sum_{j=1}^n \beta_j h_j \right)^2}{\sum_{j=1}^n \beta_j^2}. \quad (9.28)$$

При записи соотношения (9.28) учтено, что при когерентном сложении суммируются амплитуды полезных сигналов ветвей. Что же касается шумов, то в отдельных ветвях они считаются независимыми, так что суммируются их мощности (дисперсии).

Учтем теперь, что выбрано $\beta_1 = 1$ и рассмотрим пару ветвей — первую и некоторую m . Если выбрать β_m так, чтобы обеспечить максимум отношения сигнал/шум в напряжении полученном от суммирования только данной пары ветвей, то можно утверждать, что это же значение β_m будет наилучшим и с точки зрения максимизации всей суммы (9.28). Действительно, любое другое значение β_m ухудшило бы результат суммирования напряжений первой и m -й ветвей, а следовательно, и результат общего суммирования. Таким образом, в данном случае задачу оптимизации по n параметрам можно заменить задачей оптимизации только по двум из них.

Итак, пусть в (9.28) $n = 2$. Беря производную от этого соотношения по β_2 и приравнявая ее нулю, находим, что оптимальное значение веса $\beta_2 = h_2/h_1$. В соответствии с вышеизложенным это означает, что для оптимальных значений всех весов имеют место соотношения

$$\beta_j = h_j / h_1; \quad j = 1, 2, \dots, n. \quad (9.29)$$

Подставляя найденные оптимальные значения весов в (9.28), найдем максимально возможное значение h :

$$h_m^2 = \sum_{j=1}^n h_j^2 = h_{\text{в}0}^2 \sum_{j=1}^m \mu_j^2 = \chi^2 h_0^2, \quad (9.30)$$

где $\chi^2 = \frac{1}{n} \sum_{j=1}^n \mu_j^2$ — случайная величина, характеризующая результирующие замирания случайного сигнала, получающегося от когерентного сложения с оптимальными весами полезных сигналов всех ветвей разнесения, и $h_0^2 = n h_{\text{в}0}^2$. Отметим, что поскольку $\overline{\chi^2} = 1$, то

$$h_0^2 = \overline{h_m^2}. \quad (9.31)$$

Приведенные характеристики оптимального когерентного сложения были получены в предположении равенства уровней шумов в различных ветвях. Когда шумы ветвей определяются, в основном, собственными шумами их приемников, это условие выполняется непосредственно. Если же оно не выполняется, то напряжение каждой ветви должно быть перед суммированием усилено (ослаблено) таким образом, чтобы на входе сумматора указанное предположение выполнялось. Тогда можно будет использовать представленный алгоритм оптимального сложения. Отметим, что поскольку при усилении (ослаблении)

напряжения ветви соотношение сигнал/шум не меняется, соотношения (9.30) и (9.31) останутся справедливыми.

Случайная величина χ^2 представляет собой сумму квадратов независимых нормальных случайных величин, содержащую $2n$ членов. Действительно, вспомним, что величина μ^2 была рассчитана как сумма квадратов двух своих квадратур, каждая из которых предполагалась нормальной случайной величиной. По определению величины χ^2 , она представляет собой сумму квадратов всех квадратур величин μ_j . К сожалению, в общем случае плотность распределения вероятности такой случайной величины не табулирована. Рассматривается лишь частный случай, когда суммируемые нормальные случайные величины центрированы и имеют одинаковые дисперсии. В интересующем нас приложении это означает, что распределение величины χ^2 удастся записать аналитически только для случая релейевских распределений коэффициентов передачи ветвей μ_j . Можно показать, что в этом случае плотность распределения вероятностей для χ^2 принимает вид

$$P_n(x) = \frac{n}{(n-1)!} (nx)^{n-1} \exp(-xn) \text{ при } x > 0. \quad (9.32)$$

С помощью (9.32) можно вычислить среднюю вероятность ошибки на бит транслируемой последовательности в релейевском канале с медленными дружными замираниями при использовании n ветвей разнесения и когерентном сложении напряжений ветвей с оптимальными весами.

Пусть, например, используется ОФМ с некогерентной демодуляцией, так что по аналогии с (9.13) получим

$$q_{\text{ошб}} = 0,5 \exp - h_6^2 = 0,5 \exp - \chi^2 h_{60}^2. \quad (9.33)$$

Усредняя это выражение по распределению (9.32), получим, используя табулированные интегралы,

$$P_{\text{ошб}} = \overline{q_{\text{ошб}}} = \frac{1}{2} \left[\frac{n}{n + h_{60}^2} \right]^n. \quad (9.34)$$

Отметим, что величина h_{60}^2 учитывает здесь среднее значение мощности суммарного сигнала, поступившего по всем ветвям разнесения. Поэтому данную величину следует считать неизменной при сопоставлении вариантов с разным числом ветвей, ибо это будет соответствовать постоянным энергетическим ресурсам, расходуемым на передачу канальной информации. Так, при разнесении по пространству с помощью нескольких антенн это будет означать, что суммарная эффективная площадь их апертур остается постоянной независимо от величины n .

Результаты расчета по (9.34) представлены на рис. 9.6 для случаев использования двух, трех и четырех ветвей разнесения. Здесь показана для сопоставления и кривая, относящаяся к случаю отсутствия разнесения, повторяющая соответствующую зависимость на рис. 9.2. Отметим, что она полностью совпадает с зависимостью (9.33) для $n = 1$.

Рассмотрим как ведет себя вероятность ошибки при неограниченном возрастании числа ветвей разнесения, т.е. при $n \rightarrow \infty$. Переходя в (9.34) к соответствующему пределу и используя известное из математики определение числа e , согласно которому $\lim_{n \rightarrow \infty} (1 + a/n)^n = e^a$, получим

$$P_{\text{ошб}} = \frac{1}{2} \exp(-h_{60}^2), \quad (9.35)$$

что полностью совпадает с помехоустойчивостью некогерентного приема ОФМ в канале без замираний. Таким образом, оптимальный разнесенный прием способен, в принципе, полностью устранить потери энергии радиолиний, вызванные медленными замираниями.

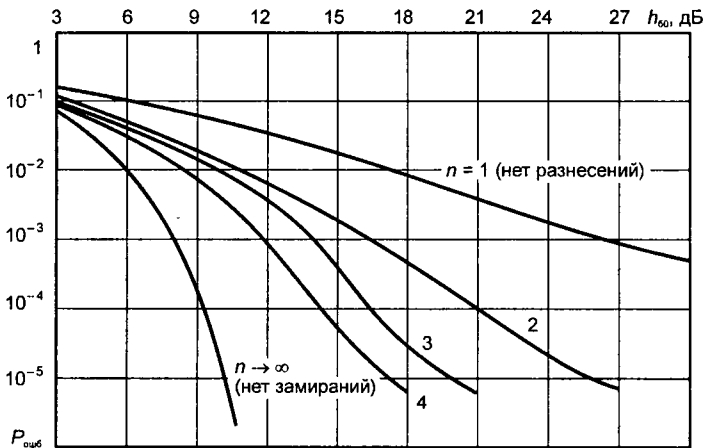


Рис. 9.6. Помехоустойчивость при разнесенном оптимальном посимвольном приеме

Данные рис. 9.6 показывают, что уже при использовании двух ветвей разнесения при когерентном весовом сложении обеспечивается значительное снижение энергетических потерь, вызванных медленными дружными замираниями. Так, для вероятности ошибки 10^{-3} эти потери снижаются с 19 до 8 дБ. Для $P_{\text{ошиб}} = 10^{-5}$ выигрыш оказывается еще большим, но, тем не менее, остаточные потери помехоустойчивости все-таки еще составляют 16 дБ. Ситуация может быть улучшена за счет увеличения числа ветвей разнесения, однако выигрыш растет сравнительно медленно, а сложность оборудования значительно нарастает. По этой причине число ветвей большее трех используется редко.

Как уже отмечалось выше, часто применяют упрощенные алгоритмы совместной обработки. Кратко остановимся на двух таких алгоритмах: квадратичного сложения и мажоритарном.

Для того чтобы пояснить идею алгоритма квадратичного сложения, вспомним структуру идеального некогерентного приемника, рассмотренную в 3.3. Решение относительно демодулированного символа в таком приемнике принимается на основе сравнения правдоподобия возможных элементарных сигналов $\lambda_{\Sigma j}$, определяемых устройствами рис. 3.7. При разнесенном приеме с квадратичным сложением блоки таких измерителей коэффициентов правдоподобия используются в каждой ветви разнесения. В результате формируется $(m \times n)$ коэффициентов правдоподобия (m — число возможных значений транслируемых символа сообщения; $j = 1, 2, \dots, m$). Обозначим коэффициенты правдоподобия через $\lambda_{\Sigma j}^k$; $k = 1, 2, \dots, n$, где k — номер ветви разнесения. При квадратичном сложении результирующие коэффициенты правдоподобия, формируются суммированием одноименных по нижнему индексу коэффициентов правдоподобия отдельных ветвей:

$$\Lambda_j = \sum_{k=1}^n \lambda_{\Sigma j}^k; \quad j = 1, 2, \dots, m. \quad (9.36)$$

Таким образом, результирующие коэффициенты правдоподобия получаются суммированием по всем ветвям коэффициентов правдоподобия для каждого возможного символа сообщения.

Описанная процедура именуется квадратурным сложением, так как в соответствии с рис. 3.7 сами величины $\lambda_{\Sigma j}$ получаются как суммы квадратов квадратурных компонент,

а следовательно, Λ_j — есть сумма таких квадратов, взятых с одноименных измерителей правдоподобия уже по всем ветвям разнесения.

После формирования результирующих коэффициентов правдоподобия, схема совместной обработки напряжений ветвей принимает решение о передававшемся символе сообщения по следующему естественному правилу: считается, что был передан тот символ, для которого Λ_j оказывается максимальным.

Квадратичное сложение существенно проще реализовать, чем оптимальное когерентное суммирование. Это следует уже из того, что оно использует гораздо меньше сведений о текущих параметрах напряжений в ветвях разнесения. Здесь не требуется знания отношений сигнал/шум и фаз сигналов ветвей. Относительно эффективности квадратичного суммирования ограничимся воспроизведением зависимостей, характеризующих его помехоустойчивость в каналах с дружными медленными релейскими замираниями, взятыми из книги [9.1] (рис. 9.7).

Сопоставляя данные рис. 9.6 и 9.7, убеждаемся, что хотя выигрыш от разнесения при оптимальном когерентном сложении больше, чем при квадратичном сложении, различие их не слишком велико (порядка 1–2 дБ).

Остановимся на мажоритарной совместной обработке. В этом варианте на каждом данном временном интервале используется сигнал только той ветви, которая по отношению сигнал/шум превосходит остальные. Если с течением времени указанное отношение в данной ветви перестает быть наибольшим, устройство совместной обработки переходит к приему сигналов другой ветви, вышедшей по этому показателю вперед. В соответствии с приведенным алгоритмом устройство мажоритарной обработки иногда именуют еще *устройством автовыбора*. Анализ, проведенный в ряде работ, показывает, что при такой обработке энергетический выигрыш от разнесения оказывается меньше, чем при квадратичном сложении на величину порядка 1 дБ.

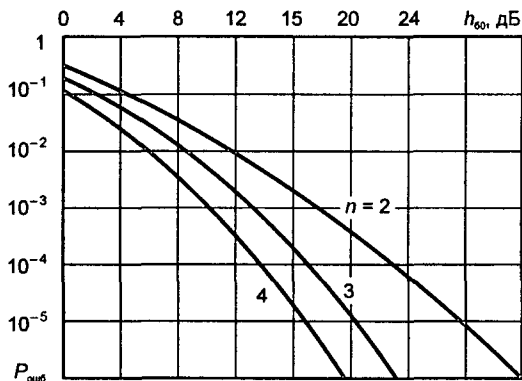


Рис. 9.7. Помехоустойчивость при разнесенном приеме с квадратичным сложением

9.3. Когерентный прием цепочек элементарных сигналов в многолучевом канале

9.3.1. Модель сигнала основной полосы

Рассмотрим простейший, но все еще практически полезный сигнал ФМ2. Комплексная огибающая последнего является вещественной функцией, которую для блока (цепочки) с N элементарными символами (ЭС) можно записать в виде

$$u(t) = \sum_{i=1}^N b_i v(t - iT_c), \quad 0 \leq t \leq NT_c,$$

где $b_i \in \{-1, +1\}$ — информационные символы; $v(t)$ — форма ЭС; T_c — длительность ЭС. В этом случае соответствующий сигнал на выходе канала с импульсной характеристикой $h(t)^*$ (на входе демодулятора) можно записать следующим образом:

$$u_{np}(t) = \int_0^t \sum_{i=1}^N b_i v(\tau - iT_c) h(t - \tau) d\tau = \sum_{i=1}^N b_i \int_0^t v(\tau - iT_c) h(t - \tau) d\tau = \sum_{i=1}^N b_i g(t - iT_c). \quad (9.37)$$

Здесь введено обозначение

$$g(t - iT_c) = \int_0^t v(\tau - iT_c) h(t - \tau) d\tau \quad (9.38)$$

для формы символа на входе демодулятора приемника, который представляет собой отклик канала на воздействие в виде ЭС.

Как правило, кроме полезного сигнала, на входе демодулятора присутствует помеха, в качестве математической модели для которой здесь примем БГШ $\xi(t)$ с математическим ожиданием, равным нулю и односторонней спектральной плотностью мощности N_0 . Тогда на входе демодулятора доступны регистрации и обработке реализации наблюдаемого процесса

$$X(t) = \sum_{i=1}^N b_i g(t - iT_c) + \xi(t), \quad 0 \leq t \leq T_n, \quad (9.39)$$

где $T_n = NT_c + T_h$ — длительность интервала наблюдения.

Введем теперь важное для дальнейшего изложения предположение: длительность отклика канала $g(t)$ ограничена и равна $T_c + T_h = T_c + mT_c$, где m — целое положительное число, т.е. вводится ограничение на память канала, которая измеряется числом m предшествующих канальных символов, влияющих на значения наблюдаемого процесса на текущем интервале времени. В этом случае $T_n = NT_c + mT_c = (N + m)T_c$ и

$$\int_0^{T_n} g(t - iT_c) g(t - jT_c) dt = \begin{cases} r_{|i-j|}, & |i-j| \leq m, \\ 0, & |i-j| > m. \end{cases} \quad (9.40)$$

Коэффициенты (9.40) будут использованы ниже при построении алгоритма демодуляции. Отметим в заключение, что если импульсная характеристика канала изменяется во времени, то изменяются и коэффициенты (9.40), в результате чего изменяется алгоритм демодуляции.

9.3.2. Оценка максимального правдоподобия последовательности информационных символов

Как уже отмечалось в гл. 3, метод максимального правдоподобия часто обеспечивает возможность построения наилучших алгоритмов выделения полезной информации из реализаций наблюдаемых процессов, являющихся различными смесями полезных сигналов и сопутствующих мешающих сторонних процессов. В этом смысле метод признается базовым при построении различных устройств выделения сигналов из помех. В том числе и в рассматриваемых в данном разделе условиях метод максимального правдоподобия позволяет указать наилучший алгоритм демодуляции сигнала на выходе многолучевого канала*.

* В этом параграфе будем рассматривать канал с неизменяющейся во времени импульсной характеристикой канала, которая к тому же предполагается известной. В последующих параграфах гл. 9 эти ограничения будут сняты.

Приведем соотношения, определяющие оценку максимального правдоподобия $\hat{\mathbf{b}} = \{\hat{b}_1, \hat{b}_2, \dots, \hat{b}_N\}$ информационной последовательности $\mathbf{b} = \{b_1, b_2, \dots, b_N\}$, вычисляемую по реализации $x(t)$ процесса $X(t)$, зарегистрированной на интервале наблюдения $0 \leq t \leq T_n$. Если информационные символы независимы и возможные их значения равновероятны, то оценка максимального правдоподобия будет обеспечивать минимальное значение вероятности ошибки приема всей последовательности. Такие оценки строятся на основе функционала отношения правдоподобия и определяются следующим соотношением [9.2]:

$$\hat{\mathbf{b}} = \{\hat{b}_1, \hat{b}_2, \dots, \hat{b}_N\} = \arg \max_{\mathbf{b}} l[x(t), 0 \leq t \leq T_n | b_1, b_2, \dots, b_N], \tag{9.41}$$

где функционал отношения правдоподобия в рассматриваемых здесь условиях

$$l[x(t), 0 \leq t \leq T_n | b_1, b_2, \dots, b_N] = \frac{\exp \left\{ -\frac{1}{N_0} \int_0^{T_n} \left[x(t) - \sum_{i=1}^N b_i g(t - iT_c) \right]^2 dt \right\}}{\exp \left\{ -\frac{1}{N_0} \int_0^{T_n} x^2(t) dt \right\}} = \tag{9.42}$$

$$= \exp \left\{ -\frac{1}{N_0} \int_0^{T_n} \left[-2x(t) \sum_{i=1}^N b_i g(t - iT_c) + \sum_{i=1}^N b_i g(t - iT_c) \sum_{j=1}^N b_j g(t - iT_c) \right] dt \right\}.$$

Запись $\hat{\mathbf{b}} = \arg \max_{\mathbf{b}} l[x(t), 0 \leq t \leq T_n | \mathbf{b}]$ условимся читать следующим образом: оценка $\hat{\mathbf{b}}$ является таким значением аргумента функционала $l[x(t), 0 \leq t \leq T_n | \mathbf{b}]$, при котором данный функционал принимает максимальное значение.

В соответствии с (9.41) и (9.42) оценка максимального правдоподобия, таким образом, обеспечивает минимальное значение целевой функции

$$Q[z_1, z_2, \dots, z_N | b_1, b_2, \dots, b_N] = -2 \sum_{i=1}^N b_i z_i + \sum_{i=1}^N \sum_{j=1}^N b_i b_j r_{|i-j|}, \tag{9.43}$$

где

$$z_i = \int_0^{T_n} x(t) g(t - iT_c) dt, \quad i = 1, 2, \dots, N \tag{9.44}$$

можно рассматривать как отсчеты мгновенных значений сигнала на выходе фильтра, согласованного с откликом канала $g(t)$. Эти отсчеты представляют собой статистики, содержащие в себе все сведения о реализации $x(t)$ процесса $X(t)$, достаточные для решения рассматриваемой здесь задачи. Второе слагаемое в правой части (9.33) характеризует наши априорные знания о свойствах канала.

Таким образом, общее определение (9.41) оценки максимального правдоподобия цепочки символов для рассматриваемых здесь условий теперь конкретизировано следующим образом:

$$\hat{\mathbf{b}} = \{\hat{b}_1, \hat{b}_2, \dots, \hat{b}_N\} = \arg \min_{\mathbf{b}} \left(-2 \sum_{i=1}^N b_i z_i + \sum_{i=1}^N \sum_{j=1}^N b_i b_j r_{|i-j|} \right). \tag{9.45}$$

В соответствии с последним выражением необходимо найти такую последовательность информационных символов $\{\hat{b}_1, \hat{b}_2, \dots, \hat{b}_N\}$, при которой целевая функция (9.43) принимает

наименьшее значение. Напомним, что каждый символ b_i принимает значения из множества $\{-1; +1\}$; аналогичным образом, каждый символ \hat{b}_i оценки (9.45) принимает значения из этого же множества*. Множество допустимых значений аргумента $\{b_1, b_2, \dots, b_N\}$ является дискретным и содержит 2^N различных значений последнего. При поиске его значения, удовлетворяющего условию (9.45), методом прямого перебора значения функции (9.43) необходимо вычислять для всех возможных значений ее аргумента с последующим сравнением получаемых значений функции. Число арифметических операций умножения и сложения, которое определяет вычислительную сложность алгоритма оценивания, нетрудно подсчитать. При $N > 100$ эта сложность оказывается столь значительной, что современные вычислительные средства не могут справиться с такой задачей за приемлемое время.

Значительное сокращение вычислительной сложности решения задачи поиска экстремума (9.45) оказывается возможным получить при использовании методов динамического программирования, сущность которого поясняется на примере одной простой задачи в следующем пункте данного раздела.

9.3.3. Пример задачи динамического программирования

Методы динамического программирования, разработанные в 1950-х гг. [9.3], к задачам связи впервые были применены при поиске эффективных алгоритмов декодирования сверточных кодов [9.4]. Затем область их применения в связи была распространена на задачи демодуляции сигналов, прошедших каналы передачи с межсимвольными искажениями, в том числе и многолучевые каналы [9.5, 9.6]. В данном разделе будет приведен пример применения динамического программирования для решения задачи поиска кратчайшего пути на некоторой решетке, которая представляет собой сеть дорог между двумя населенными пунктами. Простой численный пример позволяет проиллюстрировать как основную идею динамического программирования, так и последовательные этапы решения задачи в том виде, как это будет сделано в следующем разделе при построении алгоритма поиска экстремума целевой функции оценки максимального правдоподобия (9.45) для последовательности информационных символов.

Динамическое программирование по существу является вычислительной процедурой поиска оптимального пути или оптимальной траектории на заданной сети, имеющей ребра и узлы, пример которой приведен на рис. 9.8. Здесь шесть узлов (населенных пунктов) с номерами от 1 до 6; каждые два соседних узла (пункта) связаны ребром (дорогой); каждое ребро имеет вес (длину). Узлы 1 и 6 можно соединить различными траекториями, вес каждой из которых равен суммарному весу ребер, включенных в эту траекторию (из 1 в 6 пешеход может пройти разными дорогами, каждая из которых имеет разную длину). Как найти траекторию с минимальным весом (как найти кратчайший путь из 1 в 6)?

Очевидно, что простейшим способом решения этой задачи может служить прямой перебор: вычисление суммарных весов всех возможных траекторий (четыре траектории, восемь сложений) и выбор траектории с наименьшим весом (три сравнения). Однако данный способ не наилучший; например, метод динамического программирования обеспечивает возможность уменьшения числа вычислительных операций (сложений, сравнений). Идея динамического программирования состоит в том, что не следует вычислять сразу полные веса всех траекторий. Эти веса необходимо вычислять поэтапно.

* Подобные статистические решения в предыдущих главах назывались жесткими.

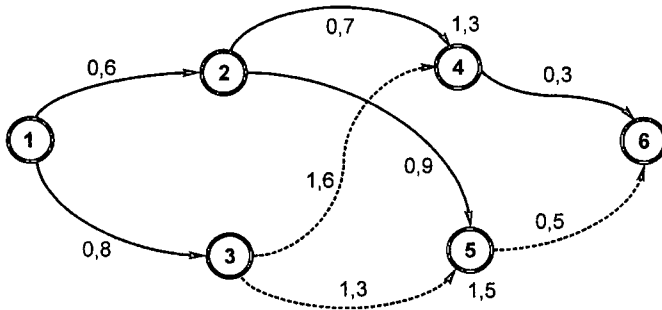


Рис. 9.8. Поиск кратчайшего пути на сети

Первый этап: вычисляются веса всех траекторий от узла 1 до узлов 4 и 5; таких траекторий четыре; для каждой вычисляется суммарный вес (всего четыре сложения); траектории до узла 4 имеют веса 1,3 и 2,4; траектории до узла 5 имеют веса 1,5 и 2,1; очевидно, что для дальнейшего анализа следует оставить траектории с наименьшими весами 1,3 до узла 4 (вес указан около узла) и с весом 1,5 до узла 5 (два сравнения); обе траектории изображены непрерывными кривыми. Остальные траектории, приходящие в эти же узлы, но имеющие большие веса, изображены штриховыми линиями; из дальнейшего анализа их можно исключить.

Второй этап: вычисляются суммарные веса траекторий от узла 4 до узла 6 (одно сложение, суммарный вес 1,6) и от узла 5 до узла 6 (одно сложение, суммарный вес 2,0); из двух траекторий выбирается та, которая имеет наименьший суммарный вес (одно сравнение, суммарный вес 1,3). Всего при таком способе поиска выполнено шесть сложений и три сравнения, вместо восьми сложений и трех сравнений при простом переборе.

Разница в числе арифметических операций будет быстро возрастать при увеличении числа этапов и увеличении числа узлов на каждом этапе. Например, если в рассматриваемой сети пришлось бы рассматривать N этапов, то число сложений оказалось бы равным $4(N-1) + 2$ при использовании метода динамического программирования, в то время как при простом переборе их будет $(N-1)2^N$, т.е. при прямом переборе число арифметических операций возрастает экспоненциально с ростом N , а при динамическом программировании — лишь линейно. При большом значении N вычислительная сложность этих алгоритмов может оказаться существенно различной. Например, в радиосистемах цифровой связи $N \geq 100$.

В рассматриваемом примере вычисления проводились от стартового узла 1 в сторону конечного узла 6 (прямой алгоритм динамического программирования). Однако эту задачу можно решать и в обратном порядке — от узла 6 в сторону узла 1 (обратный алгоритм динамического программирования).

9.3.4. Алгоритм Витерби

Покажем сначала, что поиск значения оценки максимального правдоподобия (9.45) может быть сведен к поиску некоторого пути на специальном образом построенной решетчатой диаграмме. Это позволит далее значения оценки максимального правдоподобия информационных цепочек находить с помощью алгоритмов динамического программирования.

Рассуждения начнем с построения рекуррентной записи целевой функции (9.43). Рекуррентная запись первого слагаемого данной функции имеет вид

$$2 \sum_{i=1}^N b_i z_i = 2 \sum_{i=1}^{N-1} b_i z_i + 2b_N z_N. \quad (9.46)$$

Несложно получить рекуррентную запись и второго слагаемого правой части (9.43):

$$\sum_{i=1}^N \sum_{j=1}^N b_i b_j r_{|i-j|} = \sum_{i=1}^N \left(\sum_{j=1}^{N-1} b_i b_j r_{|i-j|} + b_i b_N r_{|i-N|} \right) = \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} b_i b_j r_{|i-j|} + 2b_N \sum_{j=1}^{N-1} b_j r_{|N-j|} + b_N^2 r_0;$$

во втором слагаемом правой части этого равенства теперь необходимо учесть равенство (9.40), определяющее конечную память канала; в результате получаем

$$\sum_{i=1}^N \sum_{j=1}^N b_i b_j r_{|i-j|} = \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} b_i b_j r_{|i-j|} + 2b_N \sum_{j=N-m}^{N-1} b_j r_{|N-j|} + b_N^2 r_0. \quad (9.47)$$

Объединяя (9.46) и (9.47), имеем рекуррентную запись функции (9.43):

$$-2 \sum_{i=1}^N b_i z_i + \sum_{i=1}^N \sum_{j=1}^N b_i b_j r_{|i-j|} = -2 \sum_{i=1}^{N-1} b_i z_i + \sum_{i=1}^{N-1} \sum_{j=1}^{N-1} b_i b_j r_{|i-j|} - 2b_N z_N + 2b_N \sum_{j=N-m}^{N-1} b_j r_{|N-j|} + b_N^2 r_0. \quad (9.48)$$

Формула (9.48) записана для случая, когда известно значение целевой функции для отсчета с номером $i = N - 1$ на выходе согласованного фильтра и вычисляется приращение этой функции на последнем шаге после получения отсчета с номером $i = N$. Однако аналогичные соотношения остаются справедливыми для произвольного промежуточного шага с номером $i = k$:

$$-2 \sum_{i=1}^k b_i z_i + \sum_{i=1}^k \sum_{j=1}^k b_i b_j r_{|i-j|} = -2 \sum_{i=1}^{k-1} b_i z_i + \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} b_i b_j r_{|i-j|} - 2b_k z_k + 2b_k \sum_{j=k-m}^{k-1} b_j r_{|k-j|} + b_k^2 r_0 \quad (9.49)$$

при $k = m + 1, \dots, N$. На первом шаге при $k = m$ вычисляется начальное значение целевой функции после получения m отсчетов на выходе согласованного фильтра:

$$Q[z_1, z_2, \dots, z_m | b_1, b_2, \dots, b_m] = -2 \sum_{i=1}^m b_i z_i + \sum_{i=1}^m \sum_{j=1}^m b_i b_j r_{|i-j|}. \quad (9.50)$$

Если для приращения целевой функции на шаге с номером k ввести обозначение

$$-2b_k z_k + 2b_k \sum_{j=k-m}^{k-1} b_j r_{|k-j|} + b_k^2 r_0 = V[z_k; b_{k-m}, b_{k-m+1}, \dots, b_k], \quad (9.51)$$

то вместо (9.49) можно записать более удобное для последующего использования рекуррентное соотношение:

$$Q[z_1, z_2, \dots, z_k | b_1, b_2, \dots, b_k] = Q[z_1, z_2, \dots, z_{k-1} | b_1, b_2, \dots, b_{k-1}] + V[z_k; b_{k-m}, b_{k-m+1}, \dots, b_k] \quad (9.52)$$

при $k = m + 1, \dots, N$. Таким образом, равенство (9.52) при начальном значении (9.50) полностью определяет искомое рекуррентное соотношение для пошагового вычисления значений целевой функции в темпе поступления новых отсчетов на выходе согласованного фильтра демодулятора. Отметим также, что и начальное значение (9.50) и приращение целевой функции (9.51) зависят от текущих характеристик (9.40) многолучевого канала.

Приступим теперь к построению необходимой для использования динамического программирования решетчатой диаграммы. В соответствии с (9.40) целевая функция $Q[z_1, z_2, \dots, z_m | b_1, b_2, \dots, b_m]$ на первом шаге имеет столько начальных значений, сколько разных значений имеет последовательность $\{b_1, b_2, \dots, b_m\}$; для этой последовательности введем векторное обозначение β_m (вектор-строка). Обозначение можно сохранить и для по-

следовательностей $\{b_{k-m+1}, b_{k-m+2}, \dots, b_{k-1}, b_k\}$, если записать $\{b_{k-m+1}, b_{k-m+2}, \dots, b_{k-1}, b_k\} = \beta_k$ для значений $k = m + 1, \dots, N$. Теперь обратим внимание на то, что любой последовательности $\{b_1, b_2, \dots, b_k\}$ можно поставить во взаимнооднозначное соответствие последовательность векторов $\{\beta_m, \beta_{m+1}, \dots, \beta_k\}$; любые два соседних вектора этой последовательности имеют общие элементы. Например, при $m = 2$ и $k = 7$ цепочке

$$\{b_1, b_2, b_3, b_4, b_5, b_6, b_7\} = \{-1, -1, +1, -1, +1, +1, -1\} \tag{9.53}$$

соответствует последовательность векторов

$$\{\beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7\} = \{(-1, -1), (-1, +1), (+1, -1), (-1, +1), (+1, +1), (+1, -1)\}. \tag{9.54}$$

Использование этих обозначений позволяет записать рекуррентное соотношение (9.52) для произвольного шага с номером k следующим образом:

$$Q[z_1, z_2, \dots, z_k | \beta_m, \beta_{m+1}, \dots, \beta_k] = Q[z_1, z_2, \dots, z_{k-1} | \beta_m, \beta_{m+1}, \dots, \beta_{k-1}] + V[z_k; \beta_{k-1}, \beta_k] \tag{9.55}$$

для $k = m + 1, \dots, N$ при начальном значении (2.50) на первом шаге

$$Q[z_1, z_2, \dots, z_m | \beta_m] = -2z_m \cdot \beta_m^T + \beta_m R_m \beta_m^T, \tag{9.56}$$

где $z_m = \{z_1, z_2, \dots, z_m\}$ — вектор-строка с m компонентами, а квадратная матрица

$R_m = (r_{|i-j|})$ имеет размер $m \times m$. Последовательность $\{\beta_m, \beta_{m+1}, \dots, \beta_N\}$ будем называть последовательностью векторов состояния целевой функции $Q[z_1, z_2, \dots, z_N | \beta_m, \beta_{m+1}, \dots, \beta_N]$.

Так как между последовательностями $\{b_1, b_2, \dots, b_N\}$ и $\{\beta_m, \beta_{m+1}, \dots, \beta_N\}$ имеет место взаимнооднозначное соответствие, то вместо оценивания последовательности информационных символов можно оценивать последовательность векторов состояния $\{\beta_m, \beta_{m+1}, \dots, \beta_N\}$:

$$\{\hat{\beta}_m, \hat{\beta}_{m+1}, \dots, \hat{\beta}_N\} = \arg \min_{\beta_m, \dots, \beta_N} Q[z_1, z_2, \dots, z_N | \beta_m, \beta_{m+1}, \dots, \beta_N]. \tag{9.57}$$

Оказывается, что теперь для поиска максимального значения целевой функции в такой записи можно применять метод динамического программирования, сущность которого была изложена в 9.3.3.

Все вероятные последовательности векторов $\{\beta_m, \beta_{m+1}, \dots, \beta_N\}$ можно представить графически с помощью решетчатой диаграммы. Такая диаграмма для $m = 2$ представлена на рис. 9.9. Здесь вектор состояния $\beta_k = \{b_{k-1}, b_k\}$ может принимать значения из множества $\{(-1, -1), (+1, -1), (-1, +1), (+1, +1)\}$; эти значения указаны слева от решетки, а соответствующие им узлы решетки обозначены кружками.

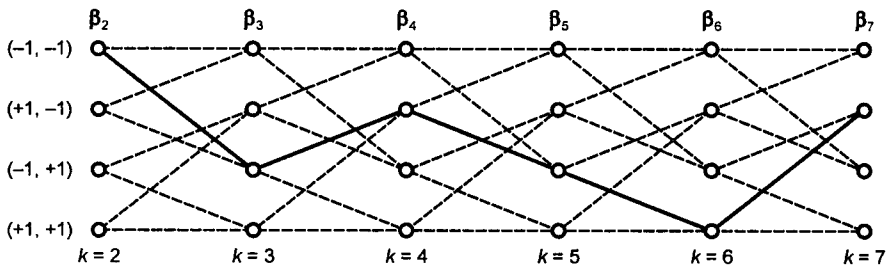


Рис. 9.9. Решетчатая диаграмма для $m = 2$

Номер шага k (указан внизу решетки) связан с вычислением по формуле (9.55) приращения и нового значения целевой функции за счет нового отсчета z_k на выходе согласованного фильтра демодулятора. Штриховые линии (далее — ребра решетки) указывают на возможные переходы от значения вектора состояния β_{k-1} на шаге с номером $k-1$ к значению вектора состояния β_k на k -м шаге (от одного узла решетки к другому узлу). Например, значению вектора состояния $\beta_4 = \{b_3, b_4\} = \{-1, +1\}$ может предшествовать лишь два из четырех возможных значений вектора $\beta_3 = \{b_2, b_3\} = \{-1, -1\}$ или $\beta_3 = \{b_2, b_3\} = \{+1, -1\}$, т.е. те значения вектора состояния $\beta_3 = \{b_2, b_3\}$, в которых информационный символ b_3 имеет то же значение, что и в векторе состояния $\beta_4 = \{b_3, b_4\}$. Конкретная последовательность векторов $\{\beta_m, \beta_{m+1}, \dots, \beta_N\}$ на решетке представляется некоторой траекторией, проходящей через соответствующие узлы. Например, последовательность (9.54) на рис. 9.9 изображена непрерывной траекторией; элементы траектории указаны над решеткой. Данной траектории соответствует последовательность информационных символов (9.53). Общее число возможных траекторий в решетке $2^m (2)^{N-m} = 2^N$, т.е. равно числу возможных информационных последовательностей.

Поиск траектории $\{\hat{\beta}_m, \hat{\beta}_{m+1}, \dots, \hat{\beta}_N\}$, удовлетворяющей условию (9.57), теперь можно осуществить с помощью алгоритма, построенного на основе динамического программирования. Формируется вектор отсчетов $z_m = \{z_1, z_2, \dots, z_m\}$ мгновенных значений сигнала на выходе согласованного фильтра демодулятора. После получения отсчета z_m (шаг $k = m$) выполняются вычисления начальных значений целевой функции по формуле (9.56) для всех 2^m возможных значений вектора состояния β_m , т.е. для всех узлов решетки. Полученные значения $Q[z_1, z_2, \dots, z_m | \beta_m] = Q[\beta_m]$ приписываются соответствующим узлам (запоминаются).

На следующем шаге с номером $k = m + 1$ после получения очередного отсчета z_{m+1} выполняются следующие вычисления:

1) для выбранного (из 2^m возможных) значения β'_{m+1} вектора состояния β_{m+1} , т.е. для выбранного узла решетки β'_{m+1} , вычисляются два приращения $V[z_{m+1}; \beta'_m, \beta'_{m+1}]$ и $V[z_{m+1}; \beta''_m, \beta'_{m+1}]$ для двух допустимых значений β'_m и β''_m вектора состояния β_m ; первое приращение приписывается ребру решетки, соединяющему узлы β'_m и β'_{m+1} , а второе — ребру, соединяющему узлы β''_m и β'_{m+1} ; оба ребра решетки входят в узел, соответствующий значению β'_{m+1} ;

2) по формуле (9.55) вычисляются два новых значения целевой функции $Q[z_1, z_2, \dots, z_{m+1} | \beta'_m, \beta'_{m+1}] = Q[z_1, z_2, \dots, z_m | \beta'_m] + V[z_{m+1}; \beta'_m, \beta'_{m+1}] = Q[\beta'_m, \beta'_{m+1}]$ и $Q[z_1, z_2, \dots, z_{m+1} | \beta''_m, \beta'_{m+1}] = Q[z_1, z_2, \dots, z_m | \beta''_m] + V[z_{m+1}; \beta''_m, \beta'_{m+1}] = Q[\beta''_m, \beta'_{m+1}]$ для двух траекторий $\{\beta'_m, \beta'_{m+1}\}$ и $\{\beta''_m, \beta'_{m+1}\}$; обе траектории входят в узел β'_{m+1} ;

3) в соответствии с критерием оптимальности (9.57) и методом динамического программирования из полученных двух накопленных значений целевой функции для дальнейшего анализа можно оставить только одно наименьшее значение $Q[\tilde{\beta}_m, \beta'_{m+1}] =$

$= \min \{ Q[\beta'_m, \beta'_{m+1}], Q[\beta''_m, \beta'_{m+1}] \}$, которое она принимает на траектории $\{\tilde{\beta}_m, \beta'_{m+1}\}$; это значение $Q[\tilde{\beta}_m, \beta'_{m+1}]$ приписывается узлу решетки β'_{m+1} ; ребро $\{\tilde{\beta}_m, \beta'_{m+1}\}$ на решетке изображается непрерывной прямой;

4) п. 1–3 выполняются для остальных $2^m - 1$ значений вектора состояния β_{m+1} , т.е. для остальных узлов решетки.

В результате вычислений на решетке из восьми возможных переходов из состояний на шаге $k = m$ в состояния на шаге $k = m + 1$ лишь четыре окажутся отмеченными сплошными линиями и будут оставлены для дальнейшего анализа. Эти траектории $\{\tilde{\beta}_m, \beta_{m+1}\}$ называют *выжившими*. Остальные траектории из последующего рассмотрения исключаются. Каждому узлу β_{m+1} будет приписано значение целевой функции $Q[\tilde{\beta}_m, \beta_{m+1}]$ на соответствующей траектории. Следует отметить, что вектор $\tilde{\beta}_m$ для выживших траекторий принимает только четыре значения, и все они рассматриваются как кандидаты, среди которых находится значение оценки максимального правдоподобия $\hat{\beta}_m$ данного вектора. Поэтому решение о значении оценки $\hat{\beta}_m$ после выполнения всех вычислений на шаге $k = m + 1$ пока принять не удастся.

На следующем шаге с номером $k = m + 2$ после получения очередного отсчета z_{m+2} выполняются все вычисления и повторяются все рассуждения, перечисленные выше в п. 1–4 предыдущего шага, если в них заменить m на $m + 1$, а $m + 1$ на $m + 2$. В результате выжившие траектории окажутся продолженными до узлов решетки β_{m+2} , которым будут приписаны значения целевой функции $Q[\tilde{\beta}_m, \tilde{\beta}_{m+1}, \beta_{m+2}]$, накопленные на шагах $k = m, k = m + 1$ и $k = m + 2$; эти значения целевая функция имеет на выживших траекториях $\{\tilde{\beta}_m, \tilde{\beta}_{m+1}, \beta_{m+2}\}$, которые на решетчатой диаграмме изображаются непрерывными ломаными линиями.

Следующий шаг вычислений выполняется аналогичным образом с очевидными изменениями в обозначениях. В результате после выполнения необходимых вычислений на шаге с номером k в узлах решетки β_k будем иметь четыре выживших траектории $\{\tilde{\beta}_m, \tilde{\beta}_{m+1}, \dots, \beta_k\}$, на которых целевая функция принимает четыре значения $Q[\tilde{\beta}_m, \tilde{\beta}_{m+1}, \dots, \beta_k]$. Следует отметить, что эти траектории на первых прошлых шагах с номерами $m, m + 1$ и т.д. до $k - L$ включительно обычно сливаются в одну, т.е. вектора $\tilde{\beta}_m, \tilde{\beta}_{m+1}, \dots, \tilde{\beta}_{k-L}$ имеют единственные значения $\hat{\beta}_m, \hat{\beta}_{m+1}, \dots, \hat{\beta}_{k-L}$, которые при выполнении вычислений на последующих шагах измениться уже не могут; следовательно, эти значения являются значениями оценок максимального правдоподобия. К сожалению, параметр L оказывается случайной величиной, и его значение зависит от конкретной реализации шума $\xi(t)$. Отметим также, что объем вычислений на очередном шаге зависит только от значения параметра m памяти канала и не зависит от количества N информационных символов в принимаемом блоке. Следовательно, вычислительная сложность рассматриваемого здесь алгоритма демодуляции является линейной функцией от длины блока N .

После завершения вычислений на последнем шаге с номером N выжившими останутся также четыре траектории $\{\tilde{\beta}_m, \tilde{\beta}_{m+1}, \dots, \tilde{\beta}_{N-1}, \beta_N\}$, на которых целевая функция имеет четыре значения $Q[\tilde{\beta}_m, \tilde{\beta}_{m+1}, \dots, \tilde{\beta}_{N-1}, \beta_N]$; в соответствии с критерием (9.57) выбора наилучшей траектории в качестве принятой демодулятором выбирается та из этих четырех, на которой целевая функция принимает наименьшее значение:

$$\{\hat{\beta}_m, \hat{\beta}_{m+1}, \dots, \hat{\beta}_{N-1}, \hat{\beta}_N\} = \arg \min_{\beta_N} Q[\hat{\beta}_m, \hat{\beta}_{m+1}, \dots, \tilde{\beta}_{N-1}, \beta_N]. \quad (9.58)$$

Оказывается, что всегда это будет единственная траектория, выделенная на решетке непрерывной ломанной прямой. В соответствии с принципами динамического программирования на траектории (9.58) достигается глобальный минимум целевой функции в задаче (9.57) и, следовательно, последовательность оценок $\{\hat{\beta}_m, \hat{\beta}_{m+1}, \dots, \hat{\beta}_{N-1}, \hat{\beta}_N\}$ однозначно определяют последовательность оценок максимального правдоподобия $\{\hat{b}_1, \hat{b}_2, \dots, \hat{b}_N\}$ для цепочки $\{b_1, b_2, \dots, b_N\}$ информационных символов переданного блока.

9.3.5. Пример реализации алгоритма Витерби

Для иллюстрации последовательности вычислений при использовании в демодуляторе алгоритма Витерби для принятия решения о переданной последовательности информационных символов примем следующие численные значения параметров канала передачи: $m = 2$, $r_0 = 1$, $r_{\pm 1} = 0,5$, $r_{\pm 2} = -0,25$; будем считать далее, что на выходе согласованного фильтра демодулятора зафиксирована выборка отсчетов $z_1 = 1,5$, $z_2 = 2,0$, $z_3 = 0,5$, $z_4 = 1,0$, $z_5 = -1,5$, $z_6 = -3,0$, $z_7 = 0,5$ [9.7]. Укажем в явном виде основные вычисления, которые предусматриваются алгоритмом Витерби при формировании оценок максимального правдоподобия $\{\hat{b}_1, \hat{b}_2, \dots, \hat{b}_7\}$ для цепочки переданных информационных символов $\{b_1, b_2, \dots, b_7\}$.

При $m = 2$ вектор состояния $\beta_2 = \{b_1, b_2\}$; аналогичным образом определяем $\beta_k = \{b_{k-1}, b_k\}$ для $k = 3, 4, 5, 6, 7$. Целевая функция (9.43) в этом примере имеет вид

$$\begin{aligned} Q[z_1, z_2, \dots, z_7 | b_1, b_2, \dots, b_7] &= Q[z_1, z_2, \dots, z_7 | \beta_2, \beta_3, \dots, \beta_7] = \\ &= -2 \sum_{i=1}^7 b_i z_i + \sum_{i=1}^7 \sum_{j=1}^7 b_i b_j r_{|i-j|}; \end{aligned} \quad (9.59)$$

в ее рекуррентной записи (9.55) приращение

$$V[z_k; b_{k-2}, b_{k-1}, b_k] = V[z_k; \beta_{k-1}, \beta_k] = -2b_k z_k + 2b_k \sum_{j=k-2}^{k-1} b_j r_{|k-j|} + b_k^2 \cdot 1 \quad (9.60)$$

при начальном значении

$$Q[z_1, z_2 | b_1, b_2] = Q[z_1, z_2 | \beta_2] = -2 \sum_{i=1}^2 b_i z_i + \sum_{i=1}^2 \sum_{j=1}^2 b_i b_j r_{|i-j|}. \quad (9.61)$$

Вектор состояния $\beta_2 = \{b_1, b_2\}$ имеет четыре разных значения, для каждого из которых по формуле (9.61) вычисляем четыре начальных значения целевой функции $Q(\beta_2)$; эти значения указаны около узлов на решетчатой диаграмме на рис. 9.10 при $k = 2$.

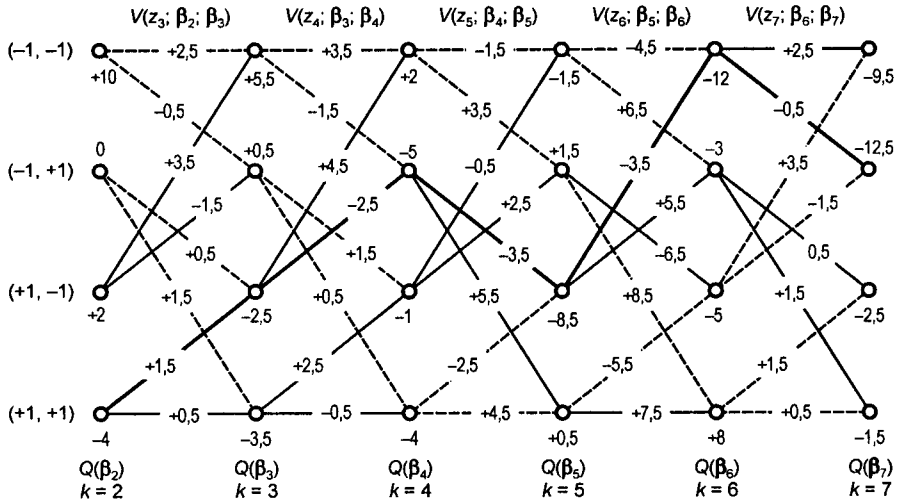


Рис. 9.10. Решетчатая диаграмма алгоритма Витерби

Далее выполняем вычисления, указанные в п. 1–4 предыдущего раздела.

Получаем значение очередного отсчета z_3 на выходе согласованного фильтра демодулятора для шага $k = 3$;

Для узла $\beta'_3 = (-1, -1)$ по формуле (9.60) вычисляем два приращения $V(z_3; \beta'_2, \beta'_3) = +2,5$ при $\beta'_2 = (-1, -1)$ и $V(z_3; \beta''_2, \beta'_3) = +3,5$ при $\beta''_2 = (-1, +1)$, которые указываем на двух ребрах, входящих в узел $\beta'_3 = (-1, -1)$.

По формуле (9.55) находим два значения целевой функции на этих ребрах: $Q[\beta'_2, \beta'_3] = 10 + 2,5 = 12,5$; $Q[\beta''_2, \beta'_3] = 2 + 3,5 = 5,5$.

В соответствии с критерием оптимальности (9.57) и принципом динамического программирования выбираем наименьшее значение $Q[\beta'_3] = Q[\beta''_2, \beta'_3] = 5,5$, которое приписываем узлу $\beta'_3 = (-1, -1)$, а ребро β''_2, β'_3 , изображенное на решетчатой диаграмме штриховой линией, заменяем непрерывной прямой.

Пункты 2, 3 выполняем для остальных трех узлов решетки при $k = 3$. Полученные значения приращений (9.60) указаны на соответствующих ребрах; выбранные значения целевой функции указаны около соответствующих узлов; выжившие траектории обозначены непрерывными прямыми.

Повторяем далее п. 1–4 последовательно для каждого шага $k = 4, 5, 6, 7$.

После выполнения вычислений на последнем шаге при $k = 7$ имеем четыре выживших траектории и четыре значения целевой функции $Q[\beta_7]$ на них; в качестве наиболее правдоподобной траектории при полученных отсчетах $\{z_1, z_2, z_3, z_4, z_5, z_6, z_7\}$ следует выбрать ту, на которой целевая функция приняла наименьшее значение; в данном примере эта траектория заканчивается в узле $\beta_7 = (-1, +1)$ и проходит через следующие узлы решетки: $\hat{\beta}_2 = (+1, +1)$, $\hat{\beta}_3 = (+1, -1)$, $\hat{\beta}_4 = (-1, +1)$, $\hat{\beta}_5 = (+1, -1)$, $\hat{\beta}_6 = (-1, -1)$, $\hat{\beta}_7 = (-1, +1)$. Соответствующая траектория на рис. 9.10 выделена утолщенной непрерывной ломаной линией.

Данной траектории соответствует следующая последовательность оценок максимального правдоподобия для информационных символов: $\hat{b}_1 = +1$, $\hat{b}_2 = +1$, $\hat{b}_3 = -1$, $\hat{b}_4 = +1$, $\hat{b}_5 = -1$, $\hat{b}_6 = -1$, $\hat{b}_7 = +1$.

Следует обратить внимание, что все траектории, выжившие после выполнения вычислений на шаге $k = 4$, проходят через один узел $\hat{\beta}_2 = (+1, +1)$. Никакие последующие значения отсчетов не изменяют этого положения. Следовательно, решения относительно значений оценок максимального правдоподобия информационных символов $\hat{b}_1 = +1$, $\hat{b}_2 = +1$ в данном примере можно было бы принять сразу после завершения вычислений на шаге $k = 4$, не ожидая поступления последующих отсчетов. На рис. 9.10 это произошло за минимальное число шагов, при котором можно из любого возможного состояния достичь любое другое состояние. Минимальное время равно памяти канала, которая в рассматриваемом примере равна $m = 2$. Однако при других значениях отсчетов и других характеристиках канала слияние выживших траекторий может не наступить вплоть до окончания обработки всего блока отсчетов.

Если канал передачи не обладает памятью, т.е. если $r_{|j-1|} = 0$ при $i \neq j$ из приведенных выше соотношений можно получить алгоритм демодуляции максимального правдоподобия для цепочек информационных символов сигнала ФМ2 с независимыми элементами, который детально был рассмотрен в гл. 3.

9.3.6. Сигналы с многопозиционной модуляцией

Комплексную огибающую такого сигнала для блока с N элементарными символами можно записать в виде

$$\dot{u}(t) = \sum_{i=1}^N \dot{b}_i v(t - iT_c), \quad 0 \leq t \leq NT_c, \quad (9.62)$$

где \dot{b}_i — комплексные информационные символы, принимающие M разных значений; $v(t)$ — форма элементарного символа; T_c — длительность ЭС. По аналогии с (9.37) комплексную огибающую сигнала на выходе канала с импульсной характеристикой $h(t) = H(t)\cos(2\pi f_0 t + \theta(t))$ (на входе демодулятора) можно записать следующим образом:

$$\dot{S}(t) = \frac{1}{2} \int_{-\infty}^t \dot{H}(t - \tau) \dot{u}(\tau) d\tau,$$

где $\dot{H}(t) = H(t)e^{j\theta(t)}$ — комплексная огибающая импульсной характеристики канала. В частности, для сигнала (9.62) на входе канала комплексная огибающая сигнала на выходе канала

$$\dot{S}(t) = \frac{1}{2} \int_0^t \sum_{i=1}^N \dot{b}_i v(\tau - iT_c) \dot{H}(t - \tau) d\tau = \sum_{i=1}^N \frac{\dot{b}_i}{2} \int_0^t v(\tau - iT_c) \dot{H}(t - \tau) d\tau = \sum_{i=1}^N \dot{d}_i \dot{g}(t - iT_c). \quad (9.63)$$

Здесь введено обозначение

$$\dot{g}(t - iT_c) = \int_0^t v(\tau - iT_c) \dot{H}(t - \tau) d\tau \quad (9.64)$$

для комплексной огибающей одного ЭС на входе демодулятора приемника.

Как и ранее будем предполагать, что длительность импульсного отклика канала ограничена интервалом $T_h = mT_c$ и $T_n = NT_c + T_h$ — длительность интервала наблюдения. Здесь

m — целое положительное число, характеризующее память канала. В этом случае $T_n = NT_c + mT_c = (N + m)T_c$ и

$$\int_0^{T_n} \dot{g}(t - iT_c) \dot{g}^*(t - lT_c) dt = \begin{cases} \dot{r}_{|i-l|}, & |i-l| \leq m, \\ 0, & |i-l| > m. \end{cases} \quad (9.65)$$

Здесь использовано обозначение \dot{x}^* для величины, комплексно сопряженной с \dot{x} .

Вместо (9.42) здесь удобно использовать комплексную запись логарифма отношения правдоподобия. Если наблюдаемый процесс на входе демодулятора является аддитивной смесью узкополосного сигнала $s(t)$ и гауссовского шума $\xi(t)$ с равномерной спектральной плотностью мощности N_0 в полосе частот сигнала, то для комплексного огибающей принимаемого сигнала справедливо представление:

$$\begin{aligned} \dot{X}(t) &= S(t)e^{j\varphi(t)} + \Xi(t)e^{j\zeta(t)} = \dot{S}(t) + \dot{\Xi}(t) \doteq S(t) \cos[\varphi(t)] + jS(t) \sin[\varphi(t)] + \\ &+ \Xi(t) \cos[\zeta(t)] + j\Xi(t) \sin[\zeta(t)] = B_s(t) + jC_s(t) + B_\xi(t) + jC_\xi(t) = \\ &= [B_s(t) + B_\xi(t)] + j[C_s(t) + C_\xi(t)] = B_x(t) + jC_x(t). \end{aligned} \quad (9.66)$$

Здесь $\dot{S}(t)$ и $\dot{\Xi}(t)$ — комплексные огибающие сигнала и шума на входе демодулятора. Теперь можно записать логарифм функционала отношения правдоподобия процесса $\dot{X}(t)$ по отношению к процессу $\dot{\Xi}(t)$ [9.2]:

$$\begin{aligned} \ln l[B_x(t), C_x(t)] &= \frac{2}{N_0} \operatorname{Re} \int_0^{T_n} \{B_x(t) + jC_x(t)\} \left\{ B_x(t) - jC_x(t) - \frac{1}{2}[B_x(t) - jC_x(t)] \right\} dt = \\ &= \frac{2}{N_0} \operatorname{Re} \int_0^{T_n} \dot{S}(t) \left\{ \dot{X}^*(t) - \frac{1}{2}[\dot{S}^*(t)] \right\} dt = \frac{2}{N_0} \operatorname{Re} \int_0^{T_n} \left\{ \dot{S}(t) \dot{X}^*(t) - \frac{1}{2}[\dot{S}(t) \dot{S}^*(t)] \right\} dt = \\ &= \ln l[\dot{X}(t)]. \end{aligned} \quad (9.67)$$

После подстановки (9.63) в (9.67) получаем окончательно

$$\begin{aligned} \ln l[\dot{X}(t)] &= \frac{2}{N_0} \operatorname{Re} \int_0^{T_n} \left\{ \sum_{i=1}^N \dot{d}_i \dot{g}(t - iT_c) \dot{X}^*(t) - \frac{1}{2} \left[\sum_{i=1}^N \dot{d}_i \dot{g}(t - iT_c) \sum_{i=1}^N \dot{d}_i \dot{g}(t - iT_c) \right] \right\} dt = \\ &= \frac{2}{N_0} \operatorname{Re} \left\{ \sum_{i=1}^N \dot{d}_i \int_0^{T_n} \dot{g}(t - iT_c) \dot{X}^*(t) dt - \frac{1}{N_0} \sum_{i=1}^N \sum_{l=1}^N \dot{d}_i \dot{d}_l^* \int_0^{T_n} \dot{g}(t - iT_c) \dot{g}^*(t - lT_c) dt \right\} = \\ &= \frac{2}{N_0} \operatorname{Re} \left\{ \sum_{i=1}^N \dot{d}_i \dot{Z}_i - \frac{1}{N_0} \sum_{i=1}^N \sum_{l=1}^N \dot{d}_i \dot{d}_l^* \dot{r}_{|i-l|} \right\} \end{aligned} \quad (9.68)$$

с аналогичными (9.44) и (9.65) обозначениями:

$$\dot{Z}_i = \int_0^{T_n} \dot{g}(t - iT_c) \dot{X}^*(t) dt. \quad (9.69)$$

Таким образом, оценка максимального правдоподобия (9.45) при многопозиционной модуляции будет определяться следующим условием:

$$\hat{\mathbf{d}} = \{ \hat{d}_1, \hat{d}_2, \dots, \hat{d}_N \} = \arg \min_{\mathbf{d}} \operatorname{Re} \left\{ - \sum_{i=1}^N \dot{d}_i \dot{Z}_i + \frac{1}{N_0} \sum_{i=1}^N \sum_{l=1}^N \dot{d}_i \dot{d}_l^* \dot{r}_{|i-l|} \right\}. \quad (9.70)$$

Каждый символ в (9.70) принимает M разных значений, так что число всех разных последовательностей длины N теперь равно M^N . Алгоритм Витерби и в данном случае существ-

венно снижает вычислительную сложность. Однако его практическая реализация оказывается заметно сложнее, чем для двоичных сигналов.

Напомним, в заключение, что в данном параграфе предполагалась известной импульсная характеристика многолучевого канала, в результате чего искаженная форма ЭС на входе демодулятора $\dot{g}(t)$ и коэффициенты (9.40) предполагаются известными. В соответствии с (9.69) и (9.70) эта функция $\dot{g}(t)$ и коэффициенты (9.40) полностью определяют оценку максимального правдоподобия цепочки информационных символов. Если характеристики канала изменяются во времени, то необходимо каким-либо способом обеспечивать адаптацию оценки (9.70). Поскольку фактическое вычисление значений этой оценки обеспечивается алгоритмом Витерби, то эта адаптация сводится к адаптации согласованного фильтра (9.69) (путем изменения опорного сигнала коррелятора) и алгоритма Витерби путем изменения значений параметров $\hat{\gamma}_{i-1|1}$, определяемых равенством (9.40). Отметим также, что представленный в данном параграфе алгоритм Витерби формирует жесткие решения.

9.4. Когерентный прием сигналов с информационной связью в многолучевом канале

9.4.1. Решетчатая диаграмма сигнала гауссовской манипуляции с минимальным сдвигом (ГММС)

Методология синтеза оптимального алгоритма приема последовательности передаваемых независимых сигналов в многолучевых каналах (см. 9.3) имеет особенность, состоящую в следующем. Из-за многолучевости канала последовательно принимаемые сигналы уже не являются независимыми, так что классические схемы оптимального приема (см. гл. 3) не могут быть применены. Таким образом, многолучевость канала привела к необходимости перехода к приему цепочек сигналов. Аналогичная ситуация возникает, если канал не многолучевой, но используются специальные методы модуляции, при которых возникает так называемая информационная связь между последовательно передаваемыми элементарными сигналами. Здесь также приходится переходить к приему цепочек сигналов и, следовательно, предусматривать меры по снижению вычислительной сложности оптимальных алгоритмов приема.

Примерами сигналов, в которых имеется связь между информационными символами, могут служить сигналы с предварительным относительным кодированием, относительной фазовой модуляцией (сотовые системы стандарта D-AMPS), с частотной манипуляцией с минимальным сдвигом и гауссовской манипуляцией с минимальным сдвигом (сотовые системы стандартов GSM-900, GSM-1800, DECT). Последние два способа модуляции детально изучены в 6.2. В данном разделе на примере этих сигналов рассматриваются вопросы когерентного приема сигналов с информационной связью в многолучевых каналах.

Из последующего изложения станет понятным, что центральной задачей здесь по-прежнему остается построение оптимального демодулятора для сигнала известной формы, которая не обязательно совпадает с формой сигнала, излучаемого передатчиком. Поскольку основным блоком демодулятора вновь оказывается устройство, реализующее алгоритм Витерби для сигнала известной формы, то в первом пункте данного параграфа особое внимание уделяется способу формирования опорных сигналов или образцов цепочек ЭС. Это позволит в дальнейшем детализировать способ адаптации алгоритма Витерби к изменяющимся характеристикам канала при частотно селективных замираниях.

Достаточно очевидно утверждение, что при наличии связи между информационными символами принятие решений в демодуляторе о каждом элементарном символе независимо от решений о соседних ЭС уже не будет наилучшим, поскольку не используется известная информация о наличии информационной связи между передаваемыми информационными символами. Поскольку в современных системах цифровой радиосвязи передача осуществляется блоками, содержащими десятки или сотни бит передаваемой информации, то даже при отсутствии причин, приводящих к возникновению межсимвольных искажений в канале передачи, задачу синтеза оптимального алгоритма демодуляции следует формулировать и решать, как задачу оптимального приема информационного блока (цепочки) в целом. Для ее решения в данном разделе вновь будет использован метод максимального правдоподобия, наиболее часто применимый к подобным задачам (он уже использовался в предыдущем разделе).

При когерентной демодуляции в приемнике любого радиосигнала сначала из принимаемого колебания (смеси сигнала и помехи на входе приемника) выделяют комплексную огибающую, из значений которой затем выделяют информационные символы. Если выбран корреляционный способ выделения (демодуляции), то в приемнике необходимо иметь возможность формировать копии возможных вариантов комплексной огибающей переданного радиосигнала на каждом тактовом интервале

$$kT_c < t \leq (k+1)T_c.$$

К сожалению, для сигнала ГММС значения сигнала на текущем интервале зависят от всех предшествующих информационных символов (см. 6.2). Тем не менее восстановить все возможные значения сигнала ГММС на очередном тактовом интервале можно, если учесть особенности его комплексной огибающей, определяемой равенствами (6.39) и (6.37). Действительно, подставляя (6.37) в (6.39) можно записать для указанного выше интервала:

$$\dot{A}(t) = A \exp \left\{ j \frac{\pi}{2} \sum_{i=-\infty}^{k-L} b_i + j \frac{\pi}{2} \sum_{i=k-L+1}^{k-1} b_i \psi(t-iT_c) + j \frac{\pi}{2} b_k \psi(t-kT_c) \right\}.$$

Первое слагаемое в показателе комплексной экспоненты представляет собой независимую от времени долю фазы комплексной огибающей на текущем интервале; второе слагаемое представляет собой функцию времени и полностью определяется $(L-1)$ предшествующими символами; последнее слагаемое есть вклад текущего информационного символа в значения фазы на данном интервале.

Очевидно, что множитель $\exp \left\{ j \frac{\pi}{2} \sum_{i=-\infty}^{k-L} b_i \right\}$ принимает значения из множества $\{1, +j, -1, -j\}$,

т.е. значение показателя этой экспоненты можно вычислять по $\text{mod } 2\pi$, так что возможные его значения принадлежат множеству $\theta = \{0, \pi/2, \pi, 3\pi/2\}$. В результате комплексную огибающую ГММС сигнала представим выражением

$$\dot{A}(t) = A \exp \{ j\theta_{k-L} \} \exp \left\{ j \frac{\pi}{2} \sum_{i=k-L+1}^{k-1} b_i \psi(t-iT_c) + j \frac{\pi}{2} b_k \psi(t-kT_c) \right\}, \quad kT_c < t \leq (k+1)T_c. \quad (9.71)$$

Выражение (9.71) можно записать в более простой и удобной форме, если ввести следующие обозначения:

$$\psi_0 = \psi(t), \quad 0 < t \leq T_c; \quad \psi_1 = \psi(t), \quad T_c < t \leq 2T_c; \quad \dots; \quad \psi_{L-1} = \psi(t), \quad (L-1)T_c < t \leq LT_c. \quad (9.72)$$

Теперь вместо (9.71) можно записать

$$\dot{A}(t) = A \exp \{ j\theta_{k-L} \} \exp \left\{ j \frac{\pi}{2} \sum_{l=1}^{L-1} b_{k-l} \psi_l + j \frac{\pi}{2} b_k \psi_0 \right\}, \quad kT_c < t \leq (k+1)T_c. \quad (9.73)$$

Представление (9.73) комплексной огибающей ГММС сигнала можно положить в основу построения функциональной схемы устройства формирования фазы этого сигнала в приемнике, которая представлена на рис. 9.11 при $L = 3^*$. Здесь регистр сдвига с элементами задержки на время T_c связан через устройства умножения с сумматором функций ψ_l , $l = 0, 1, 2$. На выходе этого сумматора формируется меняющаяся во времени доля фазы сигнала. Во втором сумматоре накапливается не меняющаяся на этом интервале доля фазы θ_{k-3} .

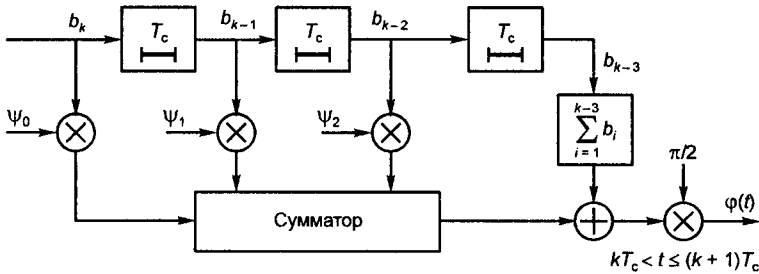


Рис. 9.11. Устройство формирования фазы ГММС сигнала при $L = 3$

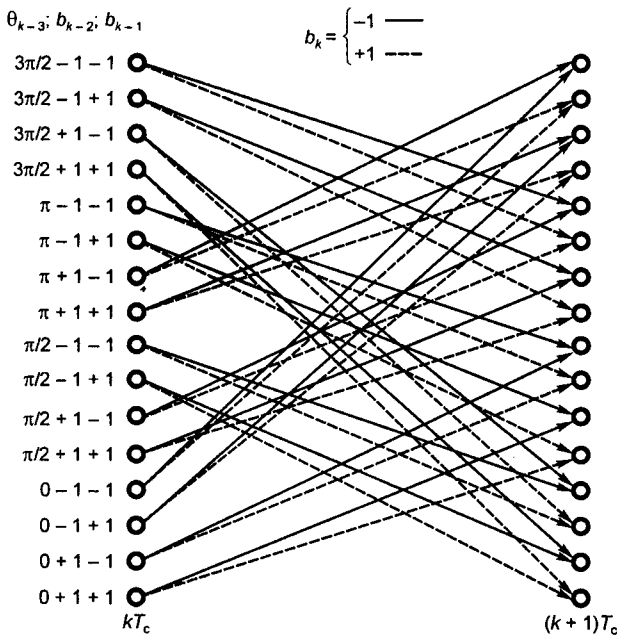


Рис. 9.12. Решетчатая диаграмма сигнала ГММС при $L = 3$

Такое представление позволяет построить решетчатую диаграмму состояний ГММС сигнала, с помощью которой сравнительно легко можно рассчитывать значения комплексной огибающей этого сигнала на любом интервале времени с номером k , если известны значения: вектора параметров состояния $[\theta_{k-3}, b_{k-2}, b_{k-1}]$ и b_k текущего символа. В установленном режиме диаграмма представляет периодическую структуру, один элемент которой представлен на рис. 9.12.

Значение вектора состояния на текущем интервале определяется информационными символами на предшествующих интервалах. При переходе от одного временного интервала к следующему наиболее старый символ включается в значение параметра θ_{k-3} и добавляется новый информационный символ.

Следовательно, из каждого состояния выходят две траектории, соответствующие двум возможным значениям -1 и $+1$ нового информационного символа, а в каждое новое состояние

* Выбрано значение, принятое в стандарте GSM-900.

входят также две траектории, соответствующие двум возможным значениям наиболее старого информационного символа.

Таким образом, на каждом интервале число разных состояний 16 (четыре начальных значения фазы θ_{k-3} и четыре разных значения двойки информационных символов $[b_{k-2}, b_{k-1}]$), а число возможных траекторий переходов равно 32. Сплошные линии соответствуют переходам, если на k -том интервале новый символ $b_k = -1$, переходы при $b_k = +1$ изображены штриховыми линиями.

Построение фазовой траектории сигнала ГММС на интервале $kT_c < t \leq (k+1)T_c$ теперь осуществляется следующим образом. Предполагаются известными значения вектора состояния $[\theta_{k-3}, b_{k-2}, b_{k-1}]$ и, следовательно, состояние слева на решетке. При появлении символа b_k его значение становится известным. Теперь значения комплексной огибающей можно вычислить по формуле (9.73) для всех моментов времени интервала с номером k . К моменту окончания данного интервала завершается переход по решетке в новое состояние, которое теперь определяется новым вектором $[\theta_{k-2}, b_{k-1}, b_k]$, где $\theta_{k-2} = \left(\theta_{k-3} + b_{k-2} \frac{\pi}{2} \right) \bmod(2\pi)$.

Решетка определяет все возможные переходы для любых вероятных значений вектора $[\theta_{k-3}, b_{k-2}, b_{k-1}]$ и двух возможных значений текущего символа b_k . В результате оказывается возможным вычислить все вероятные значения комплексной огибающей сигнала ГММС на данном интервале (для всех возможных начальных состояний). Это необходимо для формирования опорного сигнала при корреляционном способе демодуляции сигнала такого типа.

В следующем разделе данная решетчатая диаграмма будет использована при реализации алгоритма «приема сигнала в целом», так как обеспечивает возможность использования алгоритма Витерби, который обладает существенно меньшей вычислительной сложностью по сравнению с традиционным алгоритмом максимального правдоподобия.

9.4.2. Алгоритм Витерби демодуляции сигнала ГММС

В современных системах цифровой радиосвязи, где используются ГММС, информация передается блоками, каждый из которых содержит N информационных символов длительностью T_c . Прием также осуществляется блоками путем регистрации на интервале наблюдения длительностью T_n и последующей обработки наблюдаемого на входе приемника процесса $X(t)$, являющегося смесью полезного сигнала $s(t)$ и помехи $\xi(t)$. Длительность T_n интервала наблюдения несколько больше длительности $T_{\text{бл}} = NT_c$ передаваемого блока. В данном разделе будем предполагать, что помеха $\xi(t)$ представляет собой БГШ с математическим ожиданием нуль и односторонней спектральной плотностью мощности N_0 . В соответствии с (9.71) и (9.73) ГММС сигнал на интервале длительностью $T_{\text{бл}}$ имеет следующее представление:

$$s[t; u(t)] = A \cos \left[2\pi f_0 t + \frac{\pi}{2} \varphi(t) + \varphi_0 \right] = A \cos \left[2\pi f_0 t + \frac{\pi}{2} \sum_{i=0}^{N-1} b_i \psi(t - iT_c) + \varphi_0 \right]. \quad (9.74)$$

Здесь φ_0 — начальная фаза несущего колебания, которая при когерентном приеме предполагается известной; поэтому далее полагаем $\varphi_0 = 0$. Поскольку все последующее обсуждение когерентной демодуляции будет проводиться относительно одного блока ЭС, примем, что момент начала наблюдения и регистрации реализации процесса $X(t)$ совпадает с моментом начала блока, что, конечно, подразумевает наличие идеальной тактовой синхронизации; для этого момента времени примем $t_0 = 0$.

При введенных предположениях для наблюдаемого процесса на одном интервале наблюдения справедливо представление

$$X(t) = s[t; u(t)] + \xi(t), \quad 0 \leq t \leq T_n. \quad (9.75)$$

Теперь можно уточнить, что интервал наблюдения должен быть больше длительности блока T_{bn} , по крайней мере, на два символьных интервала T_c , поскольку функция $\psi(t)$ изменяется на трех символьных интервалах.

Рассматриваемая в этом разделе задача фактически сводится к построению оценки максимального правдоподобия последовательности информационных символов $\{b_i, i = 0, 1, 2, \dots, N-1\}$, каждый из них принимает значения из числового множества $\{-1, +1\}$; каждый информационный символ передается с помощью сигнала (9.74) на трех символьных интервалах; поэтому значения данного сигнала на очередном интервале с номером k длительностью T_c определяется как значением текущего информационного символа b_k , так и значениями двух предшествующих символов b_{k-2}, b_{k-1} . Более того, поскольку функция $\psi(t)$ определена на всем интервале наблюдения, то значения сигнала на k -м интервале зависят от значений и всех остальных предшествующих информационных символов принимаемого блока. В данных условиях независимая оценка каждого информационного символа невозможна и приходится искать совместные оценки всех информационных символов блока. Если для сигнала можно построить решетчатую диаграмму, то возможно использование последовательного алгоритм Витерби. Для ГММС сигнала такая решетчатая диаграмма была построена в предыдущем разделе.

Оценка максимального правдоподобия последовательности символов $\{b_i, i = 0, 1, 2, \dots, N-1\}$ рассматривалась в 9.3.2. Тот же подход сохраняется при решении этой задачи и для ГММС. Условный функционал отношения правдоподобия для наблюдаемого процесса (9.75) имеет вид

$$\begin{aligned} & l[x(t), 0 \leq t \leq T_n | b_0, b_2, \dots, b_{N-1}] = \\ & = \exp \left\{ -\frac{1}{N_0} \int_0^{T_n} \left[x(t) - A \cos[2\pi f_0 t + \frac{\pi}{2} \sum_{i=0}^{N-1} b_i \psi(t - iT_c)] \right]^2 dt \right\} / \exp \left\{ -\frac{1}{N_0} \int_0^{T_n} x^2(t) dt \right\} = \quad (9.76) \\ & = \exp \left\{ \frac{2}{N_0} \int_0^{T_n} \left[x(t) A \cos[2\pi f_0 t + \frac{\pi}{2} \sum_{i=0}^{N-1} b_i \psi(t - iT_c)] \right] dt + \frac{A^2 T_n}{2N_0} \right\}. \end{aligned}$$

Величина $\frac{A^2 T_n}{2N_0} = \frac{E_s}{N_0}$ представляет собой отношение энергии сигнала к спектральной

плотности мощности помехи $\xi(t)$ и при рассматриваемых здесь условиях передачи не зависит от значений передаваемых информационных символов. Поэтому максимизация целевой функции (9.76) путем подбора значения последовательности $\{b_i, i = 0, 1, 2, \dots, N-1\}$ фактически эквивалентна максимизации корреляционного интеграла в показателе экспоненты.

Таким образом, оценка максимального правдоподобия $\{\hat{b}_i, i = 0, 1, 2, \dots, N-1\}$ последовательности информационных символов в рассматриваемом случае должна удовлетворять следующему условию:

$$\hat{\mathbf{b}} = \{\hat{b}_0, \hat{b}_1, \dots, \hat{b}_{N-1}\} = \arg \max_{\mathbf{b}} \left(\int_0^{T_n} \left[x(t) \cos[2\pi f_0 t + \frac{\pi}{2} \sum_{i=0}^{N-1} b_i \psi(t - iT_c)] \right] dt \right). \quad (9.77)$$

Более удобное для практической реализации представление целевой функции можно получить, если воспользоваться известным разложением косинуса суммы двух углов:

$$r[x(t); \mathbf{b}] = \int_0^{T_n} \left[x(t) \cos\left[2\pi f_0 t + \frac{\pi}{2} \sum_{i=0}^{N-1} b_i \psi(t - iT_c)\right] \right] dt =$$

$$= \int_0^{T_n} x(t) \cos\left[\frac{\pi}{2} \sum_{i=0}^{N-1} b_i \psi(t - iT_c)\right] \cos(2\pi f_0 t) dt - \int_0^{T_n} x(t) \sin\left[\frac{\pi}{2} \sum_{i=0}^{N-1} b_i \psi(t - iT_c)\right] \sin(2\pi f_0 t) dt.$$

Процесс $X(t)$ является узкополосным с центральной частотой f_0 его спектральной плотности мощности; поэтому его реализация $x(t)$ представляет собой быстро осциллирующую функцию времени. Функции $\cos(2\pi f_0 t)$ и $\sin(2\pi f_0 t)$ также являются быстро осциллирующими, т.е. их период $T_0 = \frac{1}{f_0} \ll T_c$. В то же время функции $\cos\left[\frac{\pi}{2} \sum_{i=0}^{N-1} b_i \psi(t - iT_c)\right]$ и $\sin\left[\frac{\pi}{2} \sum_{i=0}^{N-1} b_i \psi(t - iT_c)\right]$ медленно изменяющиеся; можно принять, что их значения на интервале длительностью T_0 постоянны. Эти свойства подинтегральных функций можно использовать для дополнительного упрощения выражения целевой функции. Интервал интегрирования $[0, T_n]$ разобьем на последовательность малых отрезков длительностью T_0 , интеграл на интервале $[0, T_n]$ запишем в виде суммы интегралов на этих малых непересекающихся отрезках и значения медленно меняющихся функций вынесем за знаки интегралов на отрезках:

$$r[x(t); \mathbf{b}_0^{N-1}] = \sum_i \cos\left[\frac{\pi}{2} \sum_{i=0}^{N-1} b_i \psi(t_i - iT_c)\right] \int_{t_i-T_0}^{t_i} x(t) \cos(2\pi f_0 t) dt \cdot T_0 -$$

$$- \sum_i \sin\left[\frac{\pi}{2} \sum_{i=0}^{N-1} b_i \psi(t_i - iT_c)\right] \int_{t_i-T_0}^{t_i} x(t) \sin(2\pi f_0 t) dt \cdot T_0 = \tag{9.78}$$

$$= \sum_i \cos\left[\frac{\pi}{2} \sum_{i=0}^{N-1} b_i \psi(t_i - iT_c)\right] y_c(t_i) \cdot T_0 - \sum_i \sin\left[\frac{\pi}{2} \sum_{i=0}^{N-1} b_i \psi(t_i - iT_c)\right] y_s(t_i) \cdot T_0.$$

Здесь введены обозначения

$$y_c(t) = \int_{t-T_0}^{t} x(t) \cos(2\pi f_0 t) dt, \quad y_s(t) = \int_{t-T_0}^{t} x(t) \sin(2\pi f_0 t) dt \tag{9.79}$$

для значений сигналов на выходах фильтров низкой частоты с импульсной характеристикой $h(t) = 1$ при $0 \leq t \leq T_0$ и $h(t) = 0$ для всех других значений аргумента (это идеальные интеграторы с памятью T_0). Но последняя строка в (9.78) выражается разностью двух интегральных сумм, так что вместо (9.78) для целевой функции можно записать

$$r[x(t); \mathbf{b}_0^{N-1}] = \int_0^{T_n} y_c(t) \cos\left[\frac{\pi}{2} \sum_{i=0}^{N-1} b_i \psi(t - iT_c)\right] dt - \int_0^{T_n} y_s(t) \sin\left[\frac{\pi}{2} \sum_{i=0}^{N-1} b_i \psi(t - iT_c)\right] dt. \tag{9.80}$$

Как уже отмечалось в разделе 9.3, для использования алгоритма Витерби при поиске значения оценки максимального правдоподобия необходимо иметь возможность вычислять значения целевой функции последовательно по мере фиксирования реализации случайного процесса $X(t)$ на последовательности тактовых интервалов. Для получения соответствующего выражения запишем сначала целевую функцию (9.80) для случая, когда реализация $x(t)$ получена для произвольного числа k символьных интервалов:

$$r[x(t), 0 \leq t \leq kT_c; \mathbf{b}_0^{k-1}] = \int_0^{kT_c} y_c(t) \cos\left[\frac{\pi}{2} \sum_{i=0}^{k-1} b_i \psi(t - iT_c)\right] dt -$$

$$- \int_0^{kT_c} y_s(t) \sin\left[\frac{\pi}{2} \sum_{i=0}^{k-1} b_i \psi(t - iT_c)\right] dt. \tag{9.81}$$

Последующие преобразования оказываются проще, если для (9.81) использовать комплексную запись, введя следующие обозначения:

$$\dot{y}(t) = y_c(t) + jy_s(t), \quad \dot{a}(t) = \exp\left\{j\frac{\pi}{2}\sum_{i=0}^{k-1} b_i \psi(t - iT_c)\right\}$$

для комплексных огибающих принятой реализации $x(t)$ процесса $X(t)$ и опорного сигнала $s_{\text{оп}}(t)$. Вместо (9.81) можно записать

$$\begin{aligned} r[\dot{y}(t), 0 \leq t \leq kT_c; \mathbf{b}_0^{k-1}] &= \operatorname{Re} \int_0^{kT_c} \dot{y}(t) \exp\left[j\frac{\pi}{2}\sum_{i=k-3}^{k-1} b_i \psi(t - iT_c)\right] dt = \\ &= \operatorname{Re} \left\{ \int_0^{(k-1)T_c} \dot{y}(t) \exp\left\{j\frac{\pi}{2}\sum_{i=0}^{k-2} b_i \psi(t - iT_c)\right\} dt + \int_{(k-1)T_c}^{kT_c} \dot{y}(t) \exp\left\{j\frac{\pi}{2}\left[\sum_{i=0}^{k-4} b_i + \sum_{i=k-3}^{k-1} b_i \psi(t - iT_c)\right]\right\} dt \right\} = \\ &= r[\dot{y}(t), 0 \leq t \leq (k-1)T_c; \mathbf{b}_0^{k-2}] + \\ &+ \operatorname{Re} \int_{(k-1)T_c}^{kT_c} \dot{y}(t) \exp\left\{j\theta_{k-4} + j\frac{\pi}{2}\left[\sum_{i=k-3}^{k-2} b_i \psi(t - iT_c) + b_{k-1} \psi(t - (k-1)T_c)\right]\right\} dt = \\ &= r[\mathbf{b}_0^{k-2}] + \Delta r[\theta_{k-4}, \mathbf{b}_{k-3}^{k-1}]. \end{aligned} \quad (9.82)$$

Первое слагаемое в правой части этого равенства представляет собой корреляционные метрики путей (цепочек, последовательностей или векторов $\mathbf{b}_0^{k-2} = \{b_i, i = 0, 1, 2, \dots, k-2\}$) до момента времени $t_{k-1} = (k-1)T_c$; второе слагаемое определяет приращения метрик, добавляемые на интервале времени $(k-1)T_c \leq t \leq kT_c$. Заметим, что первое слагаемое зависит от всех информационных цепочек на интервале $0 \leq t \leq (k-1)T_c$, которые определяют все фазовые траектории ГММС сигнала на данном интервале; в соответствии с порядком вычислений при использовании алгоритма Витерби из данных траекторий окажутся выжившими к моменту времени $t_{k-1} = (k-1)T_c$ с соответствующими им значениями корреляционных метрик лишь незначительная часть этих траекторий. Вектор состояния сигнала к этому моменту времени имеет вид $\left\{\theta_{k-4} = \frac{\pi}{2}\sum_{i=0}^{k-4} b_i, b_{k-3}, b_{k-2}\right\}$ и может принять 16 возможных значений (четыре значения фазы θ_{k-4} и четыре значения вектора $\{b_{k-3}, b_{k-2}\}$). Следовательно, к моменту времени $t_{k-1} = (k-1)T_c$ останется только 16 выживших траекторий и, следовательно, только 16 последовательностей $\{b_i, i = 0, 1, 2, \dots, k-2\}$, которым должны быть приписаны соответствующие им 16 значений корреляционных метрик.

Второе слагаемое в правой части (9.82) в показателе экспоненты имеет выражение для всех фазовых траекторий ГММС сигнала на интервале времени $(k-1)T_c \leq t \leq kT_c$; для каждого возможного значения вектора состояния $\left\{\theta_{k-4} = \frac{\pi}{2}\sum_{i=0}^{k-4} b_i, b_{k-3}, b_{k-2}\right\}$ для данного интер-

вала возможны две фазовые траектории, соответствующие двум возможным значениям информационного символа b_{k-1} , передаваемого на этом интервале. Следовательно, на рассматриваемом интервале ГММС сигнал вновь имеет 32 возможные фазовые траектории, для каждой из которых необходимо вычислить соответствующее ей приращение корреляционной метрики. В соответствии с (9.82) последние целесообразно сложить со значениями корреляционных метрик путей (последовательностей), выживших к моменту времени $t_{k-1} = (k-1)T_c$. В результате в момент времени $t_k = kT_c$ (для $(k+1)$ -го интервала) будем иметь вектор

состояния $\left\{\theta_{k-3} = \sum_{i=0}^{k-3} b_i, b_{k-2}, b_{k-1}\right\}$, а в каждое состояние решетки в этот момент будут вхо-

дуть два пути со своими накопленными значениями корреляционных метрик. Из этих двух путей необходимо выбрать один с большим накопленным значением корреляционной метрики, которое следует приписать данному состоянию решетки. В результате к моменту времени $t_k = kT_c$ вновь останутся 16 выживших путей (цепочек, последовательностей $\{b_i, i = 0, 1, 2, \dots, k - 1\}$), для каждого из которых будет вычислено значение корреляционной метрики (9.82).

Формула (9.82) имеет рекуррентную запись. Поэтому важно указать начальное состояние всех переменных, первые и последние шаги вычислений, когда собственно должно быть принято решение о принятой последовательности информационных символов, т.е. определено значение оценки максимального правдоподобия для данной последовательности.

Порядок обработки комплексной огибающей принимаемой реализации процесса $X(t)$ здесь такой же, как и в 9.3. Регистрируются значения процесса $\dot{y}(t)$ на первых двух тактовых интервалах. Вектор состояния $\{\theta_{-1}, b_0, b_1\}$ опорного сигнала может принимать 16 различных значений. Для каждого из них вычисляются значения корреляционных метрик

$$r[\dot{y}(t), 0 \leq t \leq 2T_c; \mathbf{b}_0^1] = \text{Re} \int_0^{2T_c} \dot{y}(t) \exp \left[j\theta_{-1} + j\frac{\pi}{2} \sum_{i=0}^1 b_i \psi(t - iT_c) \right] dt = r(\mathbf{b}_0^1),$$

которые приписываются соответствующим состояниям решетки (при $k = 3$ на рис. 9.13).

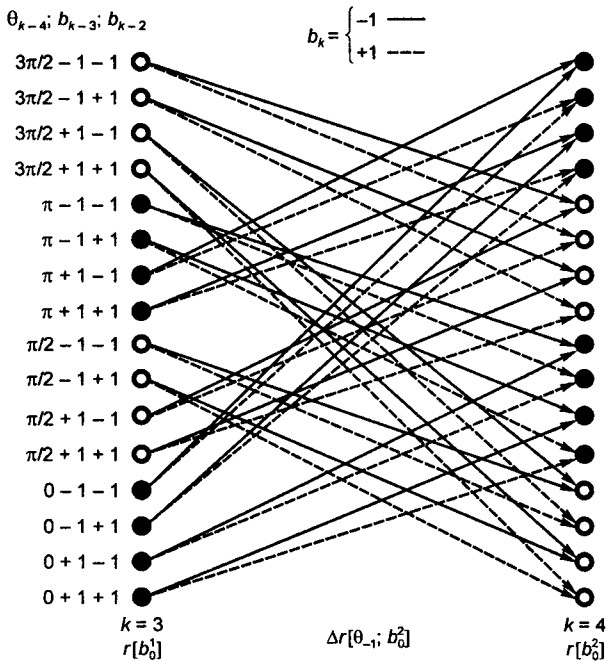


Рис. 9.13. Решетчатая диаграмма демодулятора Витерби при когерентном приеме ГММС сигнала

На третьем интервале $2T_c \leq t \leq 3T_c$ передается третий информационный символ b_2 со значениями из множества $\{-1, +1\}$. Все 16 начальных состояний и 32 возможных перехода на решетчатой диаграмме опорного сигнала (см. рис. 9.13) на этом такте изображены круж-

ками и стрелками; соответствующие им 32 фазовые траектории опорного сигнала определяются выражением

$$\Phi(t; b_0, b_1, b_2) = \theta_{-1} + \frac{\pi}{2} [b_0 \psi(t) + b_1 \psi(t - T_c) + b_2 \psi(t - 2T_c)],$$

а соответствующие им 32 приращения корреляционных метрик

$$\begin{aligned} \Delta r [\dot{y}(t), 2T_c \leq t \leq 3T_c; \theta_{-1}, \mathbf{b}_0^2] &= \int_{2T_c}^{3T_c} \dot{y}(t) \exp \left\{ j\theta_{-1} + j \frac{\pi}{2} \left[\sum_{i=0}^1 b_i \psi(t - iT_c) + b_2 \psi(t - 2T_c) \right] \right\} dt \\ &= \Delta r [\theta_{-1}, \mathbf{b}_0^2]. \end{aligned}$$

Далее для каждого нового состояния решетки при $k = 4$ находится одна выжившая траектория и соответствующее ей значение накопленной корреляционной метрики $r[\mathbf{b}_0^2]$ так, как это было описано выше в данном разделе. Аналогичные вычисления затем повторяются на четвертом интервале $3T_c \leq t \leq 4T_c$ и всех последующих вплоть до интервала $(N-1)T_c < t \leq NT_c$.

При передаче последнего информационного символа на последнем тактовом интервале соответствующая ему фазовая функция существенно изменяется еще на двух последующих интервалах. Следовательно, на этих интервалах в наблюдаемом процессе $X(t)$ содержится полезная информация о символах принимаемого информационного блока, которую желательно использовать при оценивании последних информационных символов блока. Поэтому процесс $X(t)$ следует регистрировать и обрабатывать на интервале $0 \leq t \leq NT_c + 2T_c$; длительность интервала наблюдения, следовательно, должна быть выбрана $T_H = NT_c + 2T_c$.

Таким образом, вычисления по формулам (9.82) должны выполняться до тактового интервала с номером $k = N + 1$ включительно. При этом (см. (9.7.2))

$$\begin{aligned} r[x(t), 0 \leq t \leq (N+2)T_c; \mathbf{b}_0^N] &= \\ &= \operatorname{Re} \left\{ \int_0^{NT_c} \dot{y}(t) \exp \left[j \frac{\pi}{2} \Phi(t; \mathbf{b}_0^{N-1}) \right] dt + \int_{NT_c}^{(N+1)T_c} \dot{y}(t) \exp \left[j \frac{\pi}{2} \sum_{i=0}^{N-3} b_i + j \frac{\pi}{2} \left(\sum_{i=N-2}^{N-1} b_i \psi_{N-i} + b_N \psi_0 \right) \right] dt \right\} = \\ &= \operatorname{Re} \left\{ r[\mathbf{b}_0^{N-1}] + \Delta r[\theta_{N-3}, \mathbf{b}_{N-2}^N] \right\} = \operatorname{Re} \left\{ r[\mathbf{b}_0^N] \right\} \end{aligned}$$

и

$$\begin{aligned} r[x(t), 0 \leq t \leq (N+2)T_c; \mathbf{b}_0^{N+1}] &= \\ &= \operatorname{Re} \left\{ \int_0^{(N+1)T_c} \dot{y}(t) \exp \left[j \frac{\pi}{2} \Phi(t; \mathbf{b}_0^N) \right] dt + \int_{(N+1)T_c}^{(N+2)T_c} \dot{y}(t) \exp \left[j \frac{\pi}{2} \sum_{i=0}^{N-2} b_i + j \frac{\pi}{2} \left(\sum_{i=N-1}^N b_i \psi_{N+1-i} + b_{N+1} \psi_0 \right) \right] dt \right\} = \\ &= \operatorname{Re} \left\{ r[\mathbf{b}_0^N] + \Delta r[\theta_N, \mathbf{b}_{N-1}^{N+1}] \right\} = \operatorname{Re} \left\{ r[\mathbf{b}_0^{N+1}] \right\}, \end{aligned}$$

где значения символов b_N и b_{N+1} могут быть выбраны произвольно -1 или $+1$, при условии, что эти значения известны. Например, если $b_N = b_{N+1} = -1$, то из каждого состояния решетки на такте с номером $k = N$ будет выходить лишь один путь, соответствующий значению текущего символа $b_N = -1$; таких путей будет 16, а возможных состояний решетки на последнем такте с номером $k = N + 1$ всего восемь. Поэтому после выполнения вычислений на интервале с номером $k = N$ из 16 путей выжившими останутся только восемь, а после выполнения вычислений на последнем шаге выжившими останутся только четыре пути. Из них в качестве единственного выжившего пути в соответствии с принципом максимального правдоподобия необходимо выбрать тот, который имеет наибольшее накопленное

значение корреляционной метрики $\text{Re}\{r[\mathbf{b}_0^{N+1}]\}$. Этот путь однозначно определяет значение оценки максимального правдоподобия последовательности информационных символов $\{\hat{b}_0, \hat{b}_1, \hat{b}_2, \dots, \hat{b}_{N-1}, \hat{b}_N, \hat{b}_{N+1}\}$, полученной с помощью алгоритма Витерби в результате обработки реализации $y(t)$, $0 \leq t \leq T_n$, процесса (9.75).

Основную формулу (9.82) теперь можно записать с использованием вещественных функций для $k = 4, 5, \dots, N$:

$$\begin{aligned}
 r[x(t), 0 \leq t \leq kT_c; \mathbf{b}_0^{k-1}] &= \\
 &= \int_0^{(k-1)T_c} y_c(t) \cos\left[\frac{\pi}{2}\Phi(t; \mathbf{b}_0^{k-2})\right] dt + \int_{(k-1)T_c}^{kT_c} y_c(t) \cos\left[\frac{\pi}{2}\sum_{i=0}^{k-4} b_i + \frac{\pi}{2}\left(\sum_{i=k-3}^{k-2} b_i \Psi_{k-1-i} + b_{k-1} \Psi_0\right)\right] dt = \\
 &= \int_0^{(k-1)T_c} y_s(t) \sin\left[\frac{\pi}{2}\Phi(t; \mathbf{b}_0^{k-2})\right] dt - \int_{(k-1)T_c}^{kT_c} y_s(t) \sin\left[\frac{\pi}{2}\sum_{i=0}^{k-4} b_i + \frac{\pi}{2}\left(\sum_{i=k-3}^{k-2} b_i \Psi_{k-1-i} + b_{k-1} \Psi_0\right)\right] dt. \quad (9.83)
 \end{aligned}$$

Функциональная схема части устройства когерентной демодуляции ГММС сигнала, выполняющая обработку сигнала в соответствии с формулой (9.83), изображена на рис. 9.14.

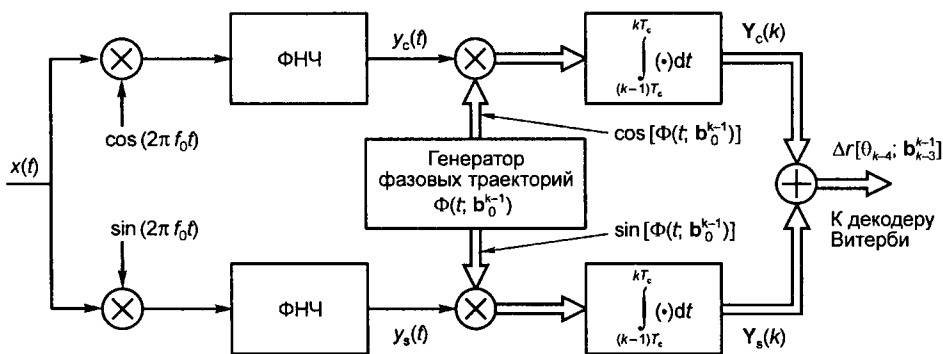


Рис. 9.14. Функциональная схема устройства вычисления приращений корреляционных метрик

Фактически на рис. 9.14 представлены элементы когерентного демодулятора, обеспечивающие вычисление приращений корреляционных метрик алгоритма Витерби на каждом шаге демодуляции. Решение о блоке информационных символов принимается после обработки всего блока.

Таким образом, информационная связь между элементарными сигналами, вводимая при передаче, даже в однолучевом канале приводит к необходимости использования приема цепочек элементарных символов, или «приема в целом».

Полученный в данном разделе алгоритм Витерби демодуляции сигнала ГММС считается основным элементом более общего адаптивного алгоритма демодуляции данного сигнала на выходе многолучевого сигнала с частотно селективными замираниями, когда форма ЭС искажается в канале и ее приходится периодически оценивать. Краткое описание такого способа борьбы с селективными замираниями, названного в 9.1 методом адаптации к селективным искажениям, приведено в следующем разделе.

9.4.3. Эквалайзер Витерби

Из предыдущего пункта следует, что если передаваемые сигналы обладают информационной связью, то в соответствии с теорией потенциальной помехоустойчивости приходится переходить к приему цепочек элементарных сигналов даже в каналах, не вносящих МСИ. Если же такие сигналы передаются по многолучевым каналам, то задача оптимального приема усложняется. Дополнительные трудности возникают в том случае, когда импульсная характеристика многолучевого канала неизвестна и меняется во времени, так что ее приходится периодически оценивать. Такой случай типичен для условий применения систем цифровой радиосвязи.

Один из методов борьбы с частотно селективными замираниями называют *выравниванием канала передачи*, понимая под ним, как правило, выравнивание передаточной функции канала, которая рассматривается как математическая модель канала в частотной области (см. 9.1). Соответствующие устройства, выполняющие необходимые для этого преобразования и называемые эквалайзерами, включаются между каналом и приемником, который при таком подходе строится в предположении отсутствия искажений сигнала в канале. Такое разделение структуры приемника на два блока приводит к некоторым потерям в качестве передачи по сравнению с приемником, структура которого построена на основе метода максимального правдоподобия с учетом как искажений сигнала в канале, так и наличия информационной связи между передаваемыми символами. Если проигрыш в качестве невелик, а снижение сложности общей структуры приемника существенное, то разделение на два независимых блока часто оказывается приемлемым техническим решением.

Иной способ борьбы с частотно селективными замираниями состоит в том, чтобы адаптировать алгоритм демодуляции к изменяющейся форме принимаемого ЭС. В этом параграфе мы приведем пример построения такого приемника для сигнала с ГММС, уделяя основное внимание вопросам адаптации алгоритма Витерби, представленного в предыдущем разделе. Коррекция искажений ЭС при таком подходе фактически не осуществляется, однако опорные сигналы алгоритма Витерби подстраиваются под оцененную искаженную форму этих сигналов. Часто такой демодулятор называют *адаптивным эквалайзером Витерби* [9.7, 9.9]. Он способен формировать как жесткие решения (см. 9.4.2), так и мягкие, о чем будет идти речь ниже в данном разделе.

Основной причиной, обуславливающей необходимость использования эквалайзера для коррекции радиоканала, является многолучевость сигнала на выходе канала (см. гл. 8). При многолучевости мощность принимаемого сигнала рассеивается во времени. Этот процесс рассеяния мощности сигнала принято характеризовать профилем задержек; наиболее важными числовыми параметрами которого являются расширение задержки Δt и среднеквадратическое рассеяние задержки σ_τ . Именно эти параметры наиболее часто используют для количественного описания искажений сигналов в многолучевых каналах. Например, такой важный параметр радиоканала, как полоса когерентности F_k , в которой амплитудно-частотная характеристика практически постоянна, а фазочастотная представляется линейной функцией частоты, часто оценивается следующим приближенным соотношением [9.8]

$$F_k \approx 1/(50\sigma_\tau).$$

Амплитуды спектральных компонент сигнала, находящихся внутри полосы когерентности, имеют коэффициент корреляции, близкий к единице. Если полосу когерентности определить как интервал между частотами двух квазигармонических колебаний, при котором огибающие имеют коэффициент корреляции 0,5, то это соотношение принимает вид

$$F_k \approx 1/(5\sigma_\tau). \quad (9.84)$$

Последние соотношения получены из статистической обработки результатов измерений. Аналитической связи между этими параметрами радиоканала, по-видимому, не существует, хотя для их пояснения можно предложить некоторые физические рассуждения. Два гармонических радиосигнала, интервал между частотами которых больше полосы когерентности, будут испытывать практически некоррелированные искажения в канале.

Например, если для городской застройки для диапазона 900 МГц среднеквадратичское расширение задержки $\sigma_\tau = 1,37$ мкс (это среднестатистическое значение для типовой застройки), то полоса когерентности (9.84) при 50%-ном уровне корреляции составит примерно

$$F_k \approx \frac{1}{5\sigma_\tau} = \frac{1}{5 \cdot 1,37 \cdot 10^{-6}} = 146 \text{ кГц.}$$

Так как полоса когерентности больше 30 кГц, то приемники сотовых систем связи стандарта AMPS при таком радиоканале могут строиться без эквалайзеров, поскольку в таком канале сигнал на входе демодулятора приемника этого стандарта будет испытывать дружные (гладкие) замирания. С другой стороны, ширина одного частотного канала системы стандарта GSM-900 равна 200 кГц и, следовательно, спектральные компоненты ГММС сигнала, частоты которых отстоят друг от друга более чем на 150 кГц, в данном канале окажутся практически некоррелированными; в канале будут иметь место частотно-селективные замирания. Приемник этого стандарта в данном канале должен иметь эквалайзер.

В цифровых системах радиосвязи демодулятор должен обеспечивать возможность формирования так называемых «мягких решений» относительно передаваемых информационных символов, которые затем подаются на вход декодера используемого помехоустойчивого кода. Это требование к демодулятору признается важным, поскольку при таком выходном сигнале демодулятора оказывается возможным получить более высокое качество передачи.

В данном разделе будет приведено краткое описание одного алгоритма функционирования эквалайзера Витерби с жесткими и мягкими решениями и способа его построения на примере цифровой сотовой системы связи стандарта GSM. Радиоинтерфейс данной системы использует диапазон частот в окрестности частоты 900 МГц и ГММС; в одном частотном канале системы организовано восемь полноскоростных каналов передачи с временным разделением, так что один физический канал представляет собой временное окно длительностью 0,577 мс и частотой повторения около 217 Гц. При скорости передачи в одном частотном канале 270,833 кб/с в одном окне передается блок, содержащий 148 закодированных информационных бит. Аналоговый речевой сигнал делится на сегменты длительностью 20 мс, которые кодируются в 50 бит класса 1а, 132 бита класса 1б и 78 бит класса 2. К битам класса 1а добавляются 3 бита проверки на четность для обнаружения ошибок в декодере. Затем биты классов 1а и 1б кодируются сверточным кодом со скоростью $\frac{1}{2}$. Выходные биты сверточного кодера и биты класса 2 перемешиваются в перемежителе и подаются на модулятор, формирующий ГММС-сигнал.

Адаптивный алгоритм демодуляции. На рис. 9.15 представлена структурная схема приемника ГММС-сигнала, в котором демодуляция искаженного каналом ГММС-сигнала осуществляется эквалайзером Витерби. Принимаемый высокочастотный сигнал когерентно детектируется с целью выделения его комплексной огибающей, затем демодулируется с помощью алгоритма Витерби, который подстраивается под искаженную каналом форму ЭС. Ошибки, возникшие в канале, обнаруживаются и корректируются в декодере канала. Полученные таким образом кодовые символы передаются в декодер речи, восстанавливающий аналоговый сигнал.

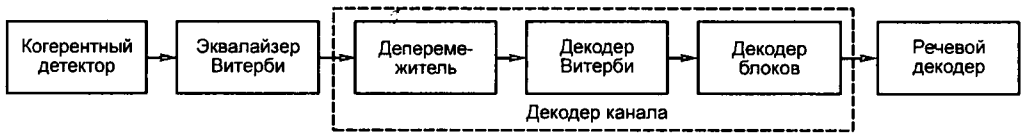


Рис. 9.15. Приемник ГММС сигнала с эквалайзером Витерби

Схема приемника будет использована для сравнительной оценки эффективности двух типов демодуляторов в виде эквалайзеров Витерби с жесткими и мягкими решениями. Применяемые в ней алгоритмы основываются на оценивании информационных символов ГММС-сигнала методом максимального правдоподобия (см. 9.4.2). Соответствующая структурная схема демодулятора изображена на рис. 9.16. Элементами данной схемы служат фильтр, согласованный с импульсной характеристикой канала, устройство оценивания импульсной характеристики канала (ИХК), блок оценивания последовательности информационных символов с алгоритмом Витерби и блок оценки достоверности формируемых оценок символов.

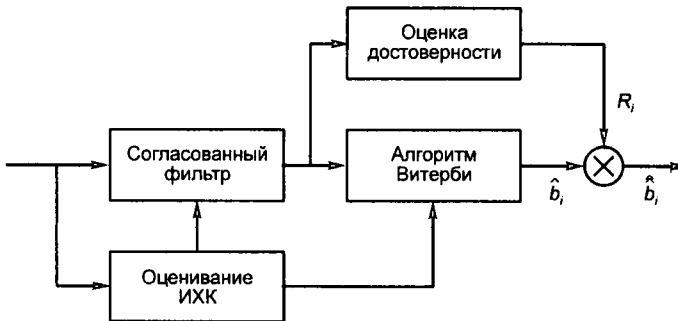


Рис. 9.16. Структурная схема эквалайзера Витерби

Импульсная характеристика канала в реальных условиях, особенно при перемещениях подвижного абонента, изменяется во времени. Однако эти изменения достаточно медленные по сравнению с изменениями мгновенных значений модулирующего сигнала на длительности одного информационного блока. Поэтому стандартом GSM предусмотрена передача обучающей последовательности в каждом передаваемом блоке данных. Такая известная последовательность имеет специальную структуру и состоит из 26 бит в середине передаваемого блока. По этой последовательности предлагается вычислять оценку ИХК для каждого принимаемого блока; предполагается, что на соседних принимаемых блоках импульсная характеристика может существенно отличаться из-за скачков по частоте, предусматриваемых стандартом GSM.

Итак, пусть импульсная характеристика канала $h(t)$ неизвестна, но доступны наблюдению, регистрации и обработке значения комплексной огибающей принимаемого сигнала $\dot{y}(t)$ на интервале наблюдения $0 \leq t \leq T_n$, в середине которого передается обучающая последовательность. Поскольку обучающая последовательность известна, то в приемнике можно сформировать на этом же интервале времени комплексную огибающую $\dot{i}(t)$ опорного сигнала, в котором модулирующим сигналом является эта же обучающая последовательность.

Покажем, что взаимную ковариационную функцию между $y(t)$ и $u(t)$ можно использовать для вычисления оценки неизвестной ИХК.

В соответствии с определением взаимная ковариационная функция между принятым и опорным сигналами при отсутствии посторонних помех может быть найдена как отклик фильтра, согласованного с опорным сигналом:

$$R_{yn}(t) = \int_{-\infty}^t y(\tau)u(t-\tau)d\tau = \int_{-\infty}^t \left[\int_{-\infty}^{\tau} u(\tau-s)h(s)ds \right] u(t-\tau)d\tau = \int_{-\infty}^t h(s) \left[\int_{-\infty}^t u(\tau-s)u(t-\tau)d\tau \right] ds = \int_{-\infty}^t h(s)R_{nn}(t-s)ds. \tag{9.85}$$

Если теперь опорный сигнал выбрать так, чтобы его ковариационная функция $R_{nn}(s)$ была близка к дельта-функции на интервале значений аргумента $-5T_c \leq s \leq 5T_c$, то на основании (9.85) можно записать

$$R_{yn}(t) = \int_{-\infty}^t h(s)R_{nn}(t-s)ds \cong h(t), \tag{9.86}$$

т.е. функцию $R_{yn}(t)$ можно рассматривать как оценку импульсной характеристики $h(t)$ на интервале $-5T_c \leq s \leq 5T_c$.

Согласно стандарту GSM цифровые системы должны проектироваться таким образом, чтобы иметь возможность компенсировать межсимвольную интерференцию на интервале времени, длительность которого равна примерно $5T_c$. Однако вычислительная сложность алгоритма обработки и сложность соответствующей аппаратной части эквалайзера Витерби удваивается, если максимальное время задержки учитываемого луча возрастает на длительность одного символа. Здесь рассмотрение ограничим интервалом расширения задержки, длительность которого равна $4T_c$. На данном интервале ограничимся только пятью отсчетами импульсной характеристики канала, отсчитываемыми последовательно в дискретные моменты времени поочередно для действительной и мнимой частей ИХК:

$$\hat{h}(t) = \sum_{i=-1}^1 h_i(t-2iT_c)\delta(t-2iT_c) + j \sum_{i=0}^1 h(t-2iT_c+T_c)\delta(t-2iT_c+T_c). \tag{9.87}$$

Фактически формула (9.87) представляет собой дискретную модель канала, представленную на рис. 9.17 и используемую при решении рассматриваемой здесь задачи. При оценивании ИХК необходимо оценивать значения вещественной и мнимой частей в указанные моменты времени. Оценка $\hat{h}(t)$ формируется на выходе фильтра, согласованного с опорным сигналом.

Импульсная характеристика согласованного фильтра демодулятора должна охватывать все существенные отсчеты ИХК. Этот фильтр адаптируется к оцениваемой импульсной характеристике канала так, чтобы выполнялось равенство:

$$\hat{h}_{сф}(t) = \hat{h}^*(-t),$$

где символ «*» обозначает комплексно сопряженную величину. Согласованный фильтр обеспечивает инвариантность демодулятора относительно сдвига символов сигнала во времени и фазы несущего колебания в когерентном детекторе.

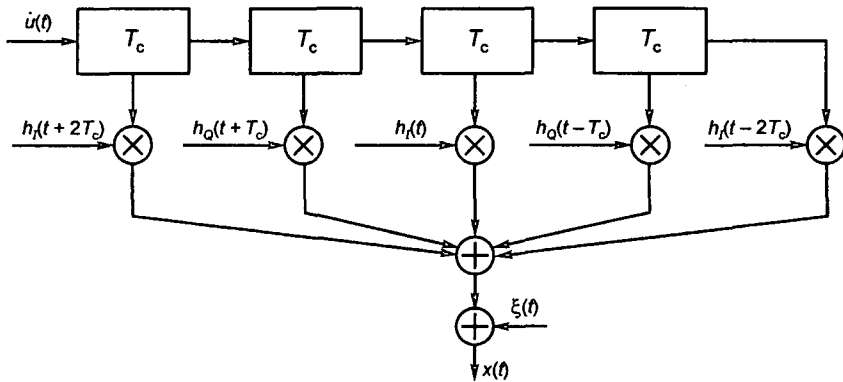


Рис. 9.17. Дискретная модель канала для комплексной огибающей сигнала

Оценки максимального правдоподобия $\hat{\mathbf{b}} = \{\hat{b}_1, \hat{b}_2, \hat{b}_3, \dots, \hat{b}_N\}$ последовательности переданных символов формируются алгоритмом Витерби (см. 9.3.4) путем последовательного вычисления приращений корреляционных метрик, которые в данном случае имеют вид (см. (9.48)):

$$r_n(\mathbf{b}_1^n) = r_{n-1}(\mathbf{b}_1^{n-1}) + \operatorname{Re} \left\{ 2b_n Y_n - b_n X_0 - 2 \sum_{m=1}^L b_{n-m} X_m \right\}.$$

Здесь

$$Y_n = Y(nT_c) = \int_{-\infty}^t \hat{y}(\tau) \hat{h}^*(t - nT_c) d\tau \quad \text{—}$$

отсчеты комплексной огибающей сигнала на выходе согласованного с каналом фильтра, получаемые с частотой дискретизации $1/T_c$,

$$X_n = X(nT_c) = \int_{-\infty}^t \hat{h}(\tau) \hat{h}^*(t - nT_c) d\tau \quad \text{—}$$

коэффициенты, характеризующие память канала; L — число бит, подвергающихся межсимвольной интерференции, влияющих на приращения корреляционных метрик путей на решетчатой диаграмме ГММС-сигнала. Таким образом, после оценивания импульсной характеристики канала, подстраивается согласованный с каналом фильтр и алгоритм Витерби.

Сформированные алгоритмом Витерби оценки $\hat{b}_i \in \{-1, +1\}$ являются жесткими решениями относительно переданных информационных символов. На основе отсчетов амплитуды сигнала на выходе согласованного с каналом фильтра вычисляются меры R_i надежности этих оценок, после чего формируются мягкие решения, выдаваемые на выход демодулятора и поступающие на вход декодера канала:

$$\hat{b}_i = R_i \cdot \hat{b}_i.$$

В блоке вычисления надежности осуществляется нормировка отсчетов сигнала с выхода согласованного фильтра.

Фактически представленный здесь эквалайзер Витерби с мягким выходом содержит в себе адаптивный алгоритм Витерби с жесткими решениями, который не подвергается каким-либо изменениям для формирования мягких решений.

Эффективность использования эквалайзера Витерби. Эффективность применения представленного выше демодулятора с адаптацией к искажениям формы ЭС в радиоканале можно оценить путем оценки вероятности ошибки воспроизведения одного бита переданной информации. Наилучшее решение данной задачи состоит в том, чтобы получить явное аналитическое выражение, связывающее вероятность ошибки с параметрами сигнала, канала и приемника. Однако это слишком трудная задача для рассматриваемого здесь приемника. Традиционный способ преодоления названной трудности состоит в статистическом имитационном моделировании исследуемой системы с применением компьютеров или проведение специально организованного эксперимента с реальным оборудованием. Здесь приведены результаты имитационного моделирования, проведенного при следующих условиях [9.10]: скорость перемещения подвижной станции 50 км/ч в городской застройке; 100 км/ч — по холмистой местности; 250 км/ч — в сельской местности. Параметры системы были выбраны в соответствии с рекомендациями стандарта GSM: сверточный кодер с длиной кодового ограничения 5, декодер Витерби; произведение длительности информационного символа на ширину полосы гауссовского фильтра $BT_c = 0,3$; исследовались эквалайзеры Витерби двух типов — с жестким и мягким выходом.

Результаты моделирования представлены на рис. 9.18 в виде графиков, иллюстрирующих зависимость оценки вероятности ошибки от отношения энергии сигнала, затрачиваемой на передачу одного бита информации, к спектральной плотности мощности N_0 шума $\xi(t)$. Графики построены для бит класса 16 (кривые 2 и 3), которые кодировались сверточным кодом, и бит

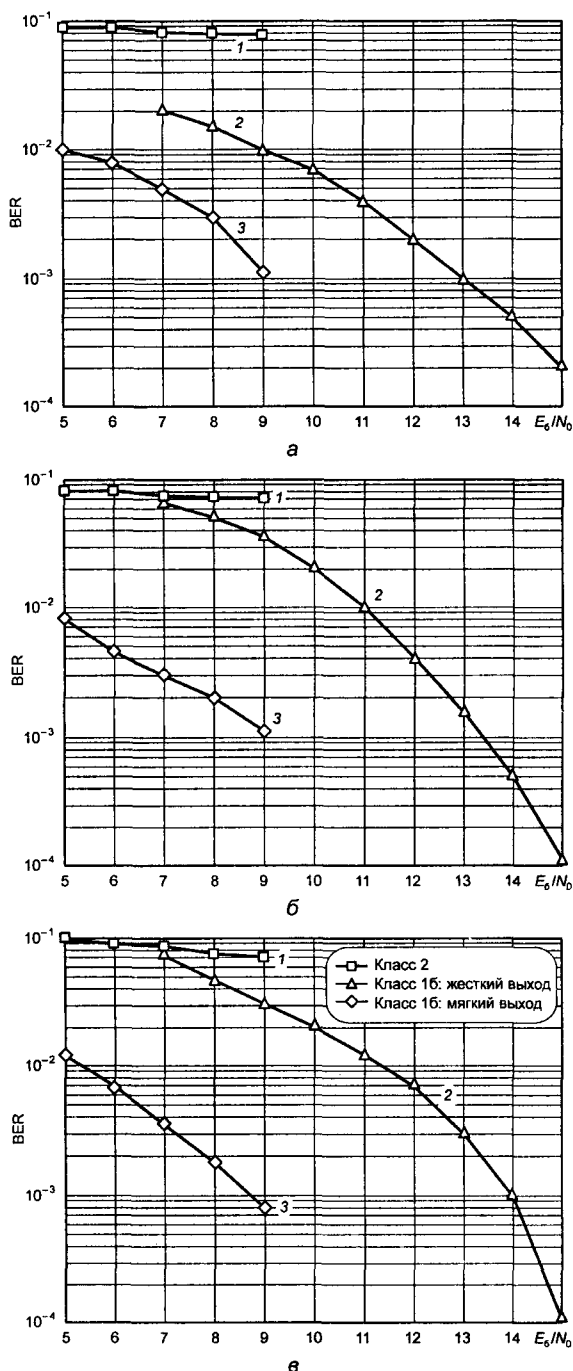


Рис. 9.18. Эффективность эквалайзера в городской застройке (а), на холмистой местности (б) и в сельской местности (в)

класса 2 (кривые 1), которые не защищались помехоустойчивым кодом. Приведенные результаты моделирования подтверждают возможность обеспечения выполнения требований стандарта GSM в рассмотренных сложных условиях распространения радиоволн при использовании эквалайзера с жестким выходом (кривые 2): требуемое значение отношения $E_b/N_0 \cong 13...14$ дБ на уровне $p_{\text{ош}} = 10^{-3}$. Приведенные графики позволяют получить количественные сравнительные оценки двух способов построения эквалайзера Витерби — с жестким выходом и мягким выходом (кривые 3). При этом следует отметить, что соответствующие структурные схемы отличаются друг от друга только наличием одного дополнительного блока вычисления коэффициента надежности. Вычислительная сложность этого блока невелика.

Однако переход к эквалайзеру с мягким выходом и, следовательно, использование декодера канала с мягким входом обеспечивают дополнительный значительный энергетический выигрыш: для обеспечения вероятности ошибки $p_{\text{ош}} = 10^{-3}$ достаточно иметь отношение $E_b/N_0 \cong 9$ дБ вместо 13...14 дБ.

На рис. 9.18 не приведены графики помехоустойчивости когерентного приема этого же сигнала без адаптивного эквалайзера Витерби, т.е. когда ИХК неизвестна и не оценивается, а прием осуществляется без учета многолучевости с помощью алгоритма, представленного в 9.3.4. Вероятность ошибки на бит передаваемой информации при рассматриваемых здесь условиях столь велика, что систему передачи приходится считать практически неработоспособной для указанных значений отношения сигнал/шум.

9.5. Ортогональное частотное разделение с мультиплексированием

9.5.1. Концепция ортогонального частотного разделения

Ортогональное частотное разделение с мультиплексированием (Orthogonal Frequency Division Multiplexing, OFDM) представляет собой специальный случай одновременной передачи потока цифровых данных по многим частотным каналам (со многими несущими или поднесущими колебаниями). Новая технология передачи в настоящее время рассматривается как одна из наиболее перспективных для построения широкополосных систем цифровой радиосвязи по многолучевым каналам, обеспечивающая достаточно высокую спектральную эффективность этих систем. Одним из привлекательных свойств данной технологии считается относительно высокая устойчивость по отношению к частотно-селективным замираниям и узкополосным помехам. В системах с одним несущим колебанием, замирание на данной частоте или узкополосная помеха, попадающая на эту частоту, могут полностью прервать передачу данных. В многочастотных системах в аналогичных условиях оказываются подавленными лишь незначительная часть поднесущих колебаний. Помехоустойчивое кодирование может обеспечить восстановление данных, потерянных на подавленных поднесущих.

При OFDM высокоскоростной поток данных разбивается на большое число низкоскоростных потоков, каждый из которых передается в своем частотном канале (на своей поднесущей частоте), т.е. в частотных каналах длительность канальных символов может быть выбрана достаточно большой, значительно превышающей время расширения задержки сигнала в канале. Следовательно, МСИ в каждом частотном канале поражает лишь незначительную часть канального символа, которую можно исключить из последующей обработки в приемнике за счет введения временного защитного интервала между соседними канальными символами при контролируемом снижении скорости передачи.

Высокая спектральная эффективность обеспечивается достаточно близким расположением частот соседних поднесущих колебаний, которые генерируются совместно так, чтобы сигналы всех поднесущих были ортогональны. Это достигается благодаря использованию дискретного преобразования Фурье (ДПФ), которое может быть эффективно выполнено с применением алгоритмов быстрого преобразования Фурье (БПФ). Следует отметить, что такое преобразование используется в приемнике данной системы передачи при демодуляции принимаемого сигнала. Благодаря этому абонентское оборудование оказывается сравнительно простым, поскольку исключается необходимость использования наборов генераторов гармонических поднесущих колебаний и когерентных демодуляторов, которые необходимы при обычном частотном разделении каналов.

Концепция использования параллельной передачи данных и частотного разделения с мультиплексированием была предложена в середине 1960-х гг. [9.11]. Особенность ее, в отличие от классического способа частотного разделения каналов, состоит в том, чтобы использовать существенно перекрывающиеся частотные каналы, в каждом из которых предлагалось организовать цифровую передачу с длительностями элементарных символов T при частотном разнесении соседних каналов на интервал $\Delta f = 1/T$. Сигналы в разных каналах оказываются ортогональными, так что межканальные помехи отсутствуют. Правда, ортогональность между сигналами возможна только в том случае, если между частотами поднесущих гармонических колебаний имеет место математически точное соотношение. Например, если на интервале T укладывается целое число периодов каждой используемой поднесущей. На рис. 9.19 представлены графики спектров одного радиоимпульса с прямоугольной огибающей и поднесущей в виде гармонического колебания с частотой f_0 и одного OFDM-символа, содержащего аналогичные радиоимпульсы на нескольких поднесущих, отстоящих по частоте друг от друга на интервалы, кратные $\Delta f = 1/T$.

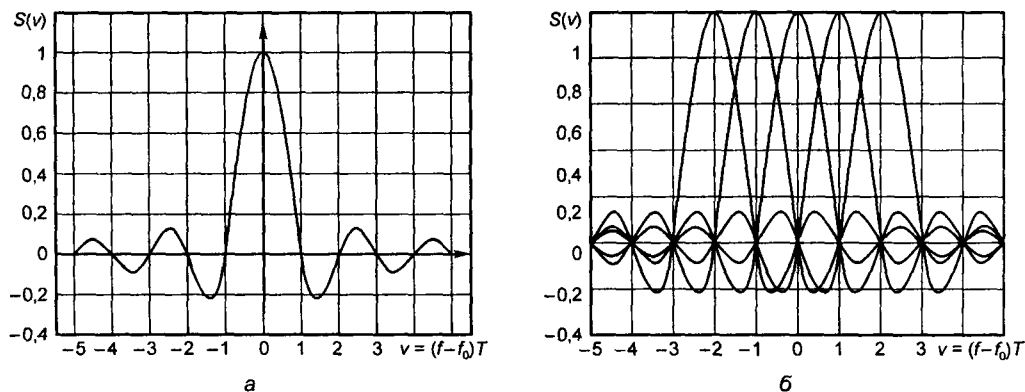


Рис. 9.19. Спектральная плотность одного элементарного символа (а) и одного OFDM-символа (б)

Технология OFDM в настоящее время используется в широкополосных цифровых системах передачи данных подвижным абонентам, высокоскоростных цифровых линиях передачи со скоростями от 1,6 до 100 Мбит/с, в цифровом радиовещании и телевидении. Основными достоинствами OFDM считаются следующие:

- в относительно медленно изменяющихся во времени каналах, в которых характеристики канала можно считать постоянными на интервале времени передачи одного

- блока данных, позволяет значительно увеличить пропускную способность посредством адаптации скорости передачи на каждой поднесущей в соответствии со значением отношения сигнал/помеха в этом частотном канале (при больших значениях отношения можно увеличивать число бит, переносимых одним элементарным символом);
- при фиксированном значении расширения задержки сложность реализации значительно ниже сложности аналогичных систем с одним несущим колебанием с эквалайзером;
 - возможность использования в сетях с одним несущим колебанием, что особенно удобно для радио- и телевидения^{*}.

С другой стороны, данной технологии присущи и некоторые недостатки:

- высокая чувствительность к смещению частоты и флюктуациям фазы принимаемого сигнала относительно опорного гармонического колебания приемника;
- относительно высокое значение отношения пиковой мощности радиосигнала к ее среднему значению, что заметно снижает энергетическую эффективность радиопередатчиков.

В настоящее время исследовательскими подразделениями различных фирм, производителей телекоммуникационного оборудования, проводятся активные исследования с целью ослабить влияние этих нежелательных свойств. Уже получены результаты, свидетельствующие о том, что без существенных потерь в качестве передачи пик-фактор OFDM-системы может не слишком сильно отличаться от значения этого показателя для системы с одним несущим колебанием и фазовой модуляцией. Получено много результатов, подтверждающих дополнительные преимущества технологии при совместном ее использовании с технологией кодового разделения каналов. Для различных приложений технология OFDM рекомендуется стандартами Международного института инженеров по электронике и электротехнике (IEEE) и Европейского института стандартизации в области телекоммуникаций (ETSI).

9.5.2. Формирование OFDM-радиосигнала

При построении OFDM-системы необходимо выбирать численные значения таких параметров, как число поднесущих гармонических колебаний, частотный интервал между соседними поднесущими, защитный интервал времени между соседними OFDM-символами, длительность канального символа, способ модуляции поднесущего колебания, методы помехоустойчивого кодирования. Значения перечисленных параметров определяются требованиями к самой системе, такими как полоса занимаемых частот, требуемая скорость передачи информации, значения расширения задержки и доплеровского расширения спектра сигнала в канале. Некоторые из этих требований противоречивы.

OFDM-сигнал представляет собой сумму поднесущих гармонических колебаний, каждая из которых модулируется своим подпотокком передаваемых бит с использованием ФМ или квадратурной амплитудной модуляции (КАМ). Пусть \dot{d}_i — комплексное число, представляющее амплитуду $|\dot{d}_i|$ и начальную фазу $\arg(\dot{d}_i)$ i -го поднесущего гармонического колебания OFDM-символа при использовании КАМ; будем называть это число КАМ-символом; каждый КАМ-символ в системах цифровой радиосвязи переносит несколько кодовых бит. Если один OFDM-символ содержит N_s поднесущих колебаний, то можно говорить, что один OFDM-символ переносит блок \dot{d}_i , $i = 0, 1, 2, \dots, N_s - 1$, КАМ-символов. Тогда ком-

^{*} Частотные каналы на поднесущих колебаниях при этой технологии формируются на видеочастоте (в окрестности нулевой частоты), а затем путем преобразования частоты переносятся в требуемую полосу радиочастот.

плексная огибающая одного OFDM-символа длительностью T , который начинается в момент времени t_k , представляется выражением

$$\dot{u}(t) = \sum_{i=0}^{N_s-1} \dot{d}_i \exp\left\{j2\pi \frac{i}{T}(t-t_k)\right\}, \quad t_k \leq t \leq t_k + T; \quad (9.88)$$

вне этого интервала времени OFDM-символ с номером k равен нулю. Формула (9.88) описывает предельный видеозэквивалент OFDM-радиосигнала. Чтобы получить реальный OFDM-символ радиосигнала с прямоугольной огибающей и частотой f_0 несущего гармонического колебания, необходимо вещественную и мнимую части огибающей (9.88), соответствующие синфазной и квадратурной компонентам КАМ-сигнала, умножить на $\cos(2\pi f_0 t)$ и $\sin(2\pi f_0 t)$ с последующим сложением полученных колебаний.

Из (9.88) следует, что для OFDM-видеосигнала интервал между частотами соседних поднесущих $\Delta f = 1/T$, частоты всех поднесущих кратны этому интервалу и, следовательно, на длительности одного OFDM-символа всегда укладывается целое число периодов каждой поднесущей. Для любых соседних поднесущих число периодов отличается на единицу. Начальная фаза и амплитуда каждой поднесущей определяются значением транслируемого этой поднесущей КАМ-символа \dot{d}_i ; для разных поднесущих они обычно оказываются разными. В таких условиях когерентная демодуляция этого сигнала может быть осуществлена с использованием взаимной ортогональности всех поднесущих на интервале $t_k \leq t \leq t_k + T$. Например, в соответствии с классической теорией потенциальной помехоустойчивости для получения оценки КАМ-символа \dot{d}_l с номером l из принятой реализации комплексной огибающей (9.88) (после ее выделения в приемнике традиционным способом) необходимо эту реализацию умножить на поднесущее колебание $\exp\left\{-j2\pi \frac{l}{T}(t-t_s)\right\}$ и результат проинтегрировать на интервале времени $t_k \leq t \leq t_k + T$:

$$\begin{aligned} & \int_{t_k}^{t_k+T} \dot{u}(t) \exp\left\{-j2\pi \frac{l}{T}(t-t_k)\right\} dt = \\ & = \int_{t_k}^{t_k+T} \exp\left\{-j2\pi \frac{l}{T}(t-t_k)\right\} \sum_{i=0}^{N_s-1} \dot{d}_i \exp\left\{j2\pi \frac{i}{T}(t-t_k)\right\} dt = \dot{d}_l T. \end{aligned} \quad (9.89)$$

Фактически в соответствии с (9.89) при демодуляции одного OFDM-символа вычисляется значение спектральной плотности амплитуд данного символа на частоте $F_l = l\Delta f$ поднесущего колебания с номером l ; но из рис. 9.19, на котором изображены перекрывающиеся спектры разных поднесущих одного OFDM-символа, видно, что спектры всех остальных поднесущих на этой частоте равны нулю. Таким образом, при таком выборе частот поднесущих и интервала между соседними частотами, связанного с длительностью символа T , в принципе можно обеспечить отсутствие взаимных влияний между поднесущими*.

Итак, комплексная огибающая одного OFDM-символа представлена выражением (9.88). Но эта формула представляет собой не что иное, как обратное преобразование Фурье совокупности КАМ-символов $\dot{d}_i, i = 0, 1, 2, \dots, N_s - 1$.

* Общее свойство идеального когерентного приемника заключается в том, что он остается идеальным и тогда, когда кроме БГШ в канале на сигнал накладываются другие сигналы, ортогональные принимаемому. На них такой приемник вообще не реагирует, поскольку он содержит коррелятор, а через него такие помехи не проходят. В частности, и здесь коррелятор, настроенный на одну поднесущую, не реагирует на остальные.

Если вместо непрерывного времени на интервале $t_k \leq t \leq t_k + T$ это преобразование вычислять только для дискретных моментов времени $t_i = t_k + i\Delta t$ и интервал дискретизации времени выбрать $\Delta t = T/(N_s - 1)$, $i = 0, 1, 2, \dots, N_s - 1$, то совокупности КАМ-символов \dot{d}_i , $i = 0, 1, 2, \dots, N_s - 1$ и отсчетов комплексной огибающей $\dot{u}(t_i)$, $i = 0, 1, 2, \dots, N_s - 1$ окажутся связанными обратным дискретным преобразованием Фурье (ОДПФ):

$$\dot{u}(t_l) = \sum_{i=0}^{N_s-1} \dot{d}_i \exp\left\{j2\pi \frac{i}{T}(t_l - t_k)\right\}, \quad l = 0, 1, 2, \dots, N_s - 1. \quad (9.90a)$$

Если принять в качестве начала отсчета времени любого OFDM-символа момент его начала, то в (9.90a) можно ограничиться только индексами:

$$\dot{u}(l) = \sum_{i=0}^{N_s-1} \dot{d}_i \exp\left\{j2\pi \frac{il}{N_s - 1}\right\}, \quad l = 0, 1, 2, \dots, N_s - 1. \quad (9.90б)$$

В настоящее время известно достаточно много эффективных алгоритмов вычисления ДПФ, из которых наибольшее распространение получили быстрые алгоритмы преобразования Фурье (БПФ). В частности, N_s — точечное обратное быстрое преобразование Фурье (ОБПФ) требует выполнения только N_s^2 комплексных умножений, которые сводятся фактически только к вращению фазы. Присутствующие операции сложения реализуются значительно проще и иногда не учитываются при сравнении вычислительной сложности различных алгоритмов. Особенно эффективны алгоритмы БПФ, если число элементов N_s может быть представлено в виде степени двойки. В этом случае число комплексных умножений при выполнении N_s -точечного ОБПФ оказывается равным $(N_s/2)\log_2(N)$ [9.12].

Таким образом, в соответствии с (9.90) формирование одного OFDM-символа радиосигнала может осуществляться следующим образом. Из потока последовательно поступающих комплексных КАМ-символов \dot{d}_i , $i = 0, 1, 2, \dots$ формируется блок, содержащий N_s символов и предназначенный для передачи одним OFDM-символом. Этот блок с помощью ОБПФ (9.90) переводится в отсчеты комплексной огибающей $\dot{u}(t_l)$, $l = 0, 1, 2, \dots, N_s - 1$. Эта операция обычно выполняется цифровыми устройствами. Сформированные таким способом цифровые отсчеты с помощью цифро-аналогового преобразователя преобразуются в аналоговую реализацию $\dot{u}(t)$, $t_k \leq t \leq t_k + T$, комплексной огибающей очередного OFDM-символа, для которой справедливо представление (9.88). Далее низкочастотный сигнал может быть перенесен на необходимую частоту несущего колебания. Однако для борьбы с многолучевостью к сформированной реализации комплексной огибающей OFDM-символа сначала добавляется так называемый префикс, способ формирования которого поясняется ниже.

9.5.3. Защитный временной интервал и циклическое продолжение

Одним из привлекательных свойств технологии OFDM является нечувствительность к расширению задержки многолучевого сигнала. Это свойство обеспечивается за счет значительного увеличения длительности одного OFDM-символа, которая больше длительности КАМ-символа в N_s раз, так что относительное время расширения задержки в канале во столько же раз уменьшается. Для того, чтобы практически полностью устранить межсимвольную интерференцию между соседними OFDM-символами, вводится защитный времен-

ной интервал T_g . Этот интервал мог бы не содержать совсем значений сигнала. Однако могла бы возникнуть межканальная интерференция между различными поднесущими, поскольку была бы нарушена их взаимная ортогональность. Причиной возникновения таких взаимных помех служит многолучевость (рис. 9.20).

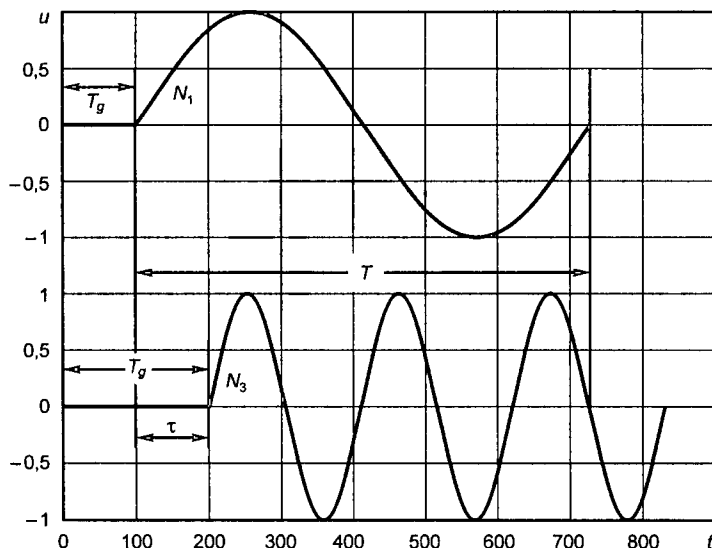


Рис. 9.20. Нарушение ортогональности поднесущих разных лучей

На рис. 9.20 изображены два разных поднесущих гармонических колебания с целым числом периодов на интервале длительностью T , каждая из которых относится к разным лучевым компонентам принимаемого OFDM-видеосигнала. Между лучами имеется временная задержка τ . Если приемник выбирает интервал интегрирования для выделения КАМ-символа поднесущей N_1 первого луча, то на этом интервале для поднесущей N_3 второго луча укладывается нецелое число периодов и, следовательно, поднесущие оказываются не ортогональными, так что поднесущая N_3 , содержащая свой КАМ-символ, внесет дополнительную мешающую поправку в результат интегрирования поднесущей N_1 . Очевидно аналогичное воздействие окажет поднесущая N_1 на результат интегрирования поднесущей N_3 на ее интервале.

Чтобы устранить нарушение ортогональности из-за многолучевости, каждое поднесущее колебание циклически продолжается на защитный интервал (рис. 9.21). В этом случае даже задержка второго луча на некоторое время τ не приведет к нарушению ортогональности между поднесущей N_1 первого луча и поднесущей N_3 второго, поскольку на любом интервале времени длительностью T всегда будет укладываться целое число периодов каждой поднесущей независимо от положения этого интервала. И это будет справедливо для любого луча, если расширение задержки Δt из-за многолучевости канала не превышает длительность защитного интервала T_g .

Время интегрирования T должно быть значительно больше длительности T_g защитного интервала. Введение защитного интервала и циклического расширения снижает скорость передачи. Сигнал на данном интервале принято называть *префиксом*. Энергия передатчика,

затрачиваемая на префикс, в приемнике не используется. Однако и энергетические потери и потери пропускной способности системы OFDM контролируются и могут быть выбраны их допустимые значения. На интервалах интегрирования каждой поднесущей в приемнике имеется сумма гармонических колебаний, амплитуды и фазы которых определяются передаваемыми ими КАМ-символами.

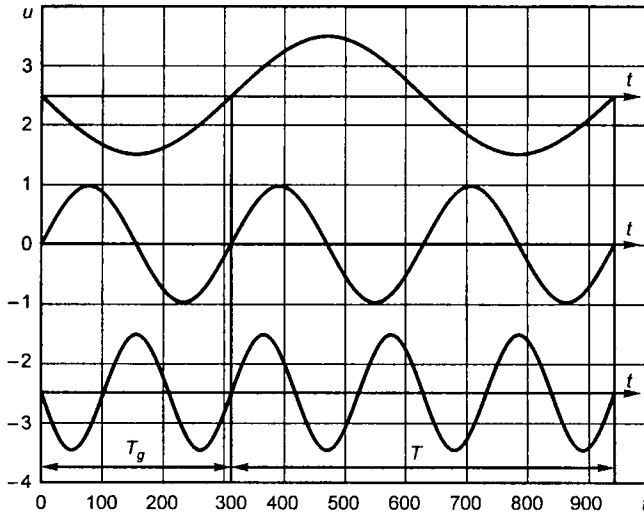


Рис. 9.21. Защитный интервал T_g и циклический префикс

Ортогональность поднесущих колебаний может нарушиться, если расширение задержки Δt в канале превышает длительность защитного интервала.

Из рис. 9.21 видно, что префикс одного OFDM-символа будет равен сумме префиксов всех поднесущих. В свою очередь, префикс одной поднесущей на защитном интервале T_g точно совпадает со значениями данной поднесущей в конце интервала интегрирования T . Поэтому префикс OFDM-видеосимвола на защитном интервале полностью совпадает со значениями комплексной огибающей этого символа в конце интервала интегрирования. Таким образом, префикс строится путем периодического продолжения мгновенных значений комплексной огибающей $\dot{u}(t)$ на защитный интервал.

9.5.4. Уменьшение внеполосных излучений

Спектральная плотность одного OFDM-символа есть сумма спектров всех поднесущих колебаний, примеры которых приведены на рис. 9.19. На рис. 9.22 показан пример графика спектра некоторой суммы поднесущих, амплитуды и фазы которых приняты одинаковыми. В реальной системе OFDM-символ представляет собой сумму модулированных КАМ-символами гармонических колебаний, каждое из которых имеет прямоугольную огибающую. В результате вне основной полосы системы цифровой радиосвязи спектр последовательности таких OFDM-символов убывает достаточно медленно, поскольку представляет собой сумму функций типа $\sin(x)/x$.

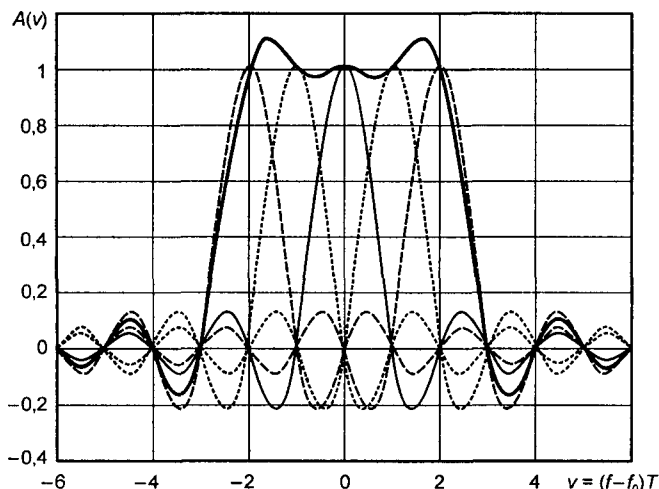


Рис. 9.22. Спектральная плотность суммы пяти поднесущих с одинаковыми амплитудами и начальными фазами

На рис. 9.23, *а* приведен пример спектральной плотности мощности OFDM-символа с 16 поднесущими. Полоса спектра на уровне минус 20 дБ уже в два раза больше ширины спектра на уровне минус 3 дБ. На рис. 9.23, *б* представлен аналогичный график для случая 128 поднесущих колебаний.

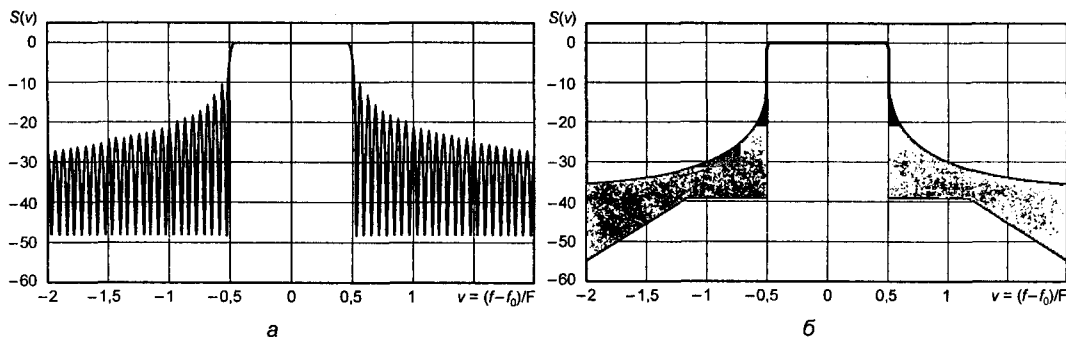


Рис. 9.23. Спектральные плотности мощности OFDM-сигнала при 16 (а) и 128 (б) поднесущих колебаний

Для уменьшения внеполосных излучений фронты комплексной огибающей каждого OFDM-символа можно сгладить. Для этого во временной области часто используется функция взвешивающего окна в виде приподнятого косинуса, которая может быть записана в следующем виде:

$$w(t) = \begin{cases} 0,5 + 0,5 \cos[\pi + \pi t / (\beta T_c)], & 0 \leq t \leq \beta T_c, \\ 1, & \beta T_c \leq t \leq T_c, \\ 0,5 + 0,5 \cos[(t - T_c)\pi / (\beta T_c)], & T_c \leq t \leq (1 + \beta)T_c. \end{cases} \quad (9.91)$$

Здесь $T + T_g = T_c$, β называется коэффициентом скругления; длительность одного OFDM-символа $(1 + \beta)T_c$ на величину βT_c оказывается больше T_c , а соседние OFDM-символы немного перекаываются (рис. 9.24).

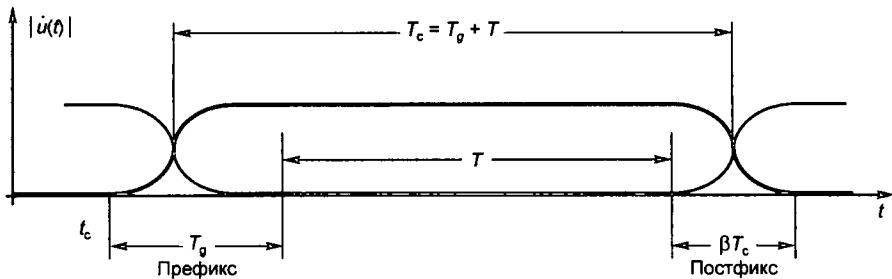


Рис. 9.24. Параметры вещественной огибающей OFDM-символа

Теперь вместо (9.90) можно записать общее выражение, определяющее комплексную огибающую OFDM-символа с учетом защитного интервала, префикса и сглаживающего временного окна:

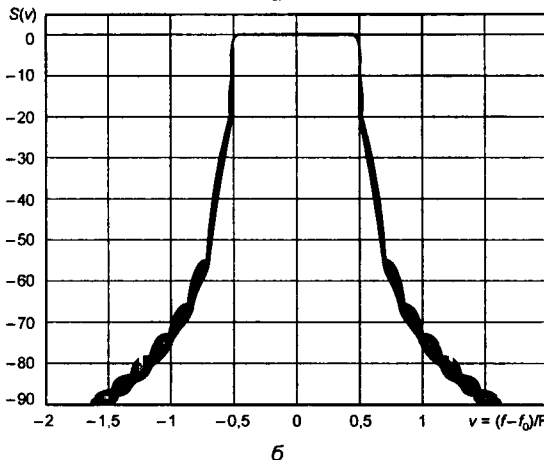
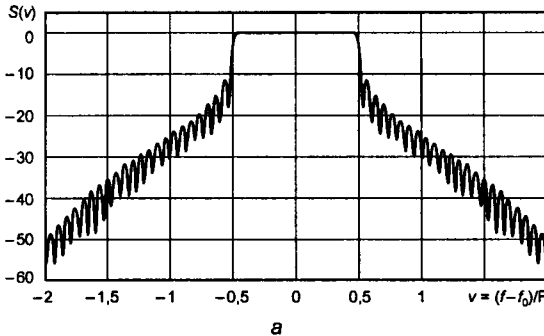


Рис. 9.25. Спектральные плотности мощности OFDM-сигнала при 16 (а) и 128 (б) поднесущих колебаний для коэффициента скругления $\beta = 0,05$

$$\begin{aligned} \dot{y}(t_l) &= w(t_l - t_c) \sum_{i=0}^{N_s-1} \dot{d}_i \times \\ &\times \exp \left\{ j2\pi \frac{i}{T} (t_l - t_c - T_g) \right\}, \\ t_c &\leq t_l \leq t_c + (1 + \beta)T_c. \end{aligned} \quad (9.91)$$

Здесь отсчеты комплексной огибающей OFDM-символа берутся в моменты времени $t_l = t_c + l \frac{T}{N_s - 1}$, т.е. шаг

дискретизации по времени $\Delta t = T/(N_s - 1)$ оставлен прежним. Практически отсчеты OFDM-символа формируются следующим образом: после выполнения ОБПФ все последние отсчеты полученного массива, оказавшиеся на интервале длительностью T_g , вставляются в начало массива (на место защитного интервала, часть которого будет использована для скругления переднего фронта OFDM-символа) и часто называются *отсчетами префикса*, а все первые отсчеты массива ОБПФ, оказавшиеся на интервале времени длительностью βT_c , добавляются в конец массива ОБПФ (на место скругления огибающей заднего фронта) и именуются *отсчетами пост-*

фикса. После такого дополнения массива ОБПФ получается массив отсчетов комплексной огибающей OFDM-видеосимвола, который умножается на весовое окно.

Спектр произведения весового окна на значения комплексной огибающей OFDM-символа можно найти как свертку спектра весового окна с набором дельта-функций, расположенных на частотах поднесущих. На рис. 9.25 представлены графики спектров OFDM-символов с огибающей (9.91) для $\beta = 0,05$ при числе поднесущих 16 и 128, которые можно сравнить с аналогичными спектрами при прямоугольных огибающих, изображенными на рис. 9.23.

На рис. 9.26 приведена соответствующая всем перечисленным выше соображениям структурная схема OFDM-передатчика.

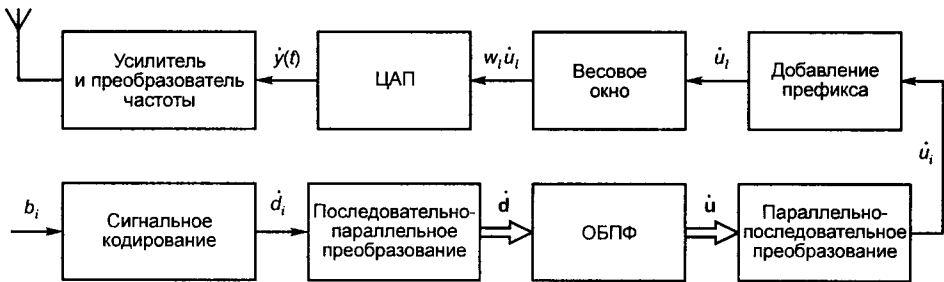


Рис. 9.26. Структурная схема OFDM-передатчика

К основным параметрам схемы на рис. 9.26 относятся полоса частот F , предоставляемая линии передачи, скорость передачи R (число кодовых бит в секунду), расширение задержки Δt сигнала в многолучевом канале. Расширение задержки непосредственно определяет длительность защитного интервала T_g : защитный интервал должен быть больше в 2–4 раза среднеквадратического значения расширения задержки σ_t . Выбор конкретного значения зависит от способа модуляции и типа помехоустойчивого кодирования. При более высокой значности модуляции (КАМ64) приходится выбирать большее значение данного множителя, поскольку такой сигнал более чувствителен к межсимвольной интерференции, возникающей из-за многолучевости. Более сложное кодирование ослабляет эту зависимость и значение коэффициента можно выбрать меньшим.

После выбора защитного интервала можно найти длительность OFDM-символа T_c . Чем меньшую долю интервала T_c составляет длительность защитного интервала T_g , тем меньше энергетические потери и, следовательно, меньше требуемая излучаемая мощность для обеспечения необходимого значения отношения сигнал/помеха в приемнике. Однако длительность T_c не может быть слишком большой, так как число поднесущих колебаний становится слишком большим, и частотные интервалы между ними слишком малыми. Возрастает чувствительность системы к отклонениям частоты в приемнике и к флуктуациям фазы. Возрастает также пикфактор OFDM-сигнала. Последнее затрудняет обеспечение линейного режима мощных усилителей передатчика. Практическая рекомендация по выбору длительности T_c символа может быть сформулирована следующим образом: если допустить энергетические потери не более 1 дБ, то длительность символа должна быть больше защитного интервала по крайней мере в пять раз. Следовательно, длительность основного интервала интегрирования OFDM-символа в приемнике $T = T_c - T_g$, а частотный интервал между соседними поднесущими колебаниями $\Delta f = 1/T$.

Теперь можно определить число поднесущих $N_s = F/\Delta f$. С другой стороны, число поднесущих можно определить как отношение суммарной скорости передачи линии R к скорости передачи R_n на одной поднесущей. В свою очередь, скорость передачи на одной поднесущей определяется видом модуляции (например, КАМ16), скоростью используемого кода, скоростью канальных символов. При найденной скорости следования OFDM-символов и заданной скорости передачи линии можно вычислить число бит, переносимых одним символом и т.д.

9.5.5. Когерентная демодуляция OFDM-сигнала

При когерентной демодуляции OFDM-сигнала в приемнике необходимо точно знать опорные фазы и амплитуды всех поднесущих колебаний, что обеспечивает возможность установки границ областей на сигнальном созвездии для принятия решений о КАМ-символах. Такая информация, как правило, отсутствует, так что эти фазы и амплитуды приходится оценивать по принимаемому колебанию. В результате сигнальное созвездие каждой поднесущей в приемнике оказывается искаженным — изменены значения фаз и амплитуд из-за неточности частотной и тактовой синхронизации и частотно-селективных замираний.

На рис. 9.27 представлена структурная схема когерентного приемника. Полагаем, что в приемнике обеспечена идеальная частотная, фазовая и тактовая синхронизация. В этом случае точно известны моменты начала OFDM-символов. После выделения комплексной огибающей $\hat{y}(t)$ принимаемого сигнала с помощью АЦП формируется последовательность отсчетов $\hat{y}(t_i)$, $t_c \leq t_i \leq t_c + (1+\beta)T_c$. Все отсчеты, соответствующие защитному интервалу T_g (префиксу), отбрасываются; для дальнейшей обработки выделяется только N_s отсчетов, расположенных на основном интервале интегрирования T . Отсчеты, поступающие последовательно во времени, накапливаются в устройстве, выполняющем последовательно-параллельное преобразование, на выход которого выдается вектор \hat{y} с компонентами $\hat{y}(t_i)$, $i = 0, 1, 2, \dots, N_s - 1$.

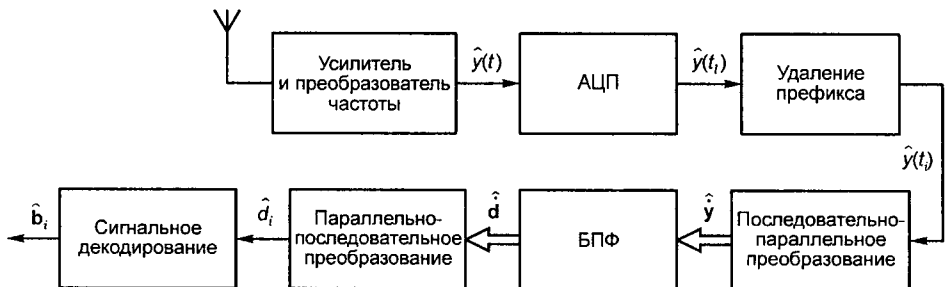


Рис. 9.27. Когерентный приемник OFDM-сигнала

Основная операция демодуляции — выделение или оценивание КАМ-символов — выполняется с помощью БПФ, посредством которого формируется оценка $\hat{\mathbf{d}}$ вектора \mathbf{d} . Компоненты оценки $\hat{\mathbf{d}}$ последовательно передаются в блок сигнального декодирования. Представленный здесь алгоритм приема следует несколько уточнить.

Действительно, многолучевой канал (см. гл. 8) можно характеризовать импульсной характеристикой $\hat{h}(t)$. Поэтому вектор отсчетов \hat{y} на выходе канала (на входе приемника) представляется сверткой сигнала и импульсной характеристики канала. Поскольку в блоке БПФ вычисляется преобразование Фурье этого вектора, то для компоненты \hat{d}_i справедливо представление:

$$\hat{d}_i = \dot{H}_i \dot{d}_i + \eta_i, \quad i = 0, 1, 2, \dots, N_s - 1, \quad (9.93)$$

где η_i — случайная величина, являющаяся преобразованием Фурье от аддитивного шума на входе приемника, а

$$\dot{H}_i = \sum_{l=0}^{N_s-1} \dot{h}(l \frac{T}{N_s-1}) \exp \left\{ -j2\pi \frac{l \cdot i}{N_s-1} \right\}, \quad i = 0, 1, 2, \dots, N_s - 1 \quad (9.94)$$

представляет собой коэффициент передачи канала на частоте поднесущей с номером i . Поскольку коэффициент \dot{H}_i — комплексная величина, то КАМ-символ \hat{d}_i отличается от переданного символа \dot{d}_i и по амплитуде, и по фазе, даже при отсутствии шума. Поэтому до передачи символа \hat{d}_i в блок сигнального декодирования канальный множитель \dot{H}_i необходимо оценить и скомпенсировать.

Для оценки канального множителя \dot{H}_i используется «обучающий» сигнал либо в виде известной модулирующей последовательности на каждой поднесущей, либо в виде немодулированной поднесущей. Если параметры канала изменяются во времени сравнительно медленно, то обучающий сигнал может вставляться поочередно в последовательности передаваемых КАМ-символов каждой поднесущей или по заранее определенному алгоритму их очередности. Формируемые при этом оценки канальных коэффициентов будут отслеживать изменения импульсной характеристики канала, так что система OFDM становится адаптивной.

При появлении глубоких замираний на частоте поднесущей с номером i , когда $|\dot{H}_i|$ оказывается значительно ниже своего среднего значения, в данном частотном канале возможно значительное увеличение вероятности ошибки. Для борьбы с такими замираниями могут быть использованы методы помехоустойчивого кодирования с разнесением по частоте, которые рассматривались в 7.5.2 для систем с широкополосными сигналами и псевдослучайной прыгающей рабочей частотой.

Глава 10

СИНХРОНИЗАЦИЯ В СИСТЕМАХ ЦИФРОВОЙ РАДИОСВЯЗИ

10.1. Общие данные о системах автоматического регулирования, используемых для синхронизации

10.1.1. Общие сведения

В предыдущих разделах неоднократно отмечалась необходимость использования в системах цифровой радиосвязи различных систем синхронизации. Так, теория идеального приема сигналов (см. гл. 3) основана на наличии в приемнике данных о положении тактовых моментов цифрового сообщения, для получения которых должны использоваться системы тактовой синхронизации (СТС). Кроме того, идеальный прием требует априорного знания частоты несущей принимаемого сигнала, а когерентный его вариант — еще и ее фазы. Эти параметры определяются с помощью систем синхронизации по несущей (СНН) приемника. В гл. 6 в общих чертах была описана еще одна система синхронизации приемника — система цикловой синхронизации (СЦС). Необходимость использования специальной системы синхронизации — системы синхронизации по задержке (ССЗ) — при обработке широкополосных сигналов в приемнике отмечалось в гл. 7.

Перечисленные системы целиком замыкаются в приемнике. Часто приходится распределять элементы системы синхронизации между различными комплексами радиолинии. Это относится, например, к также упоминавшимся в предыдущих разделах системам синхронизации пакетов (ССП). В сотовой связи стандарта GSM они реализуются в абонентских терминалах с использованием данных о текущих ошибках синхронизации, поступающих на эти терминалы с базовых станций (БС).

В качестве источника информации (опорных сигналов — ОС) для работы систем синхронизации могут использоваться как сам принимаемый модулированный информационный радиосигнал, так и специально передаваемые по радиолинии служебные сигналы. Достоинство первого варианта — отсутствие потерь пропускной способности канала на передачу ОС. Однако работа систем синхронизации резко усложняется наличием манипуляции сигнала заранее неизвестным на приемной стороне информационным сообщением. При этом часто манипуляции подвергается тот же параметр сигнала, который должен восстанавливаться системой синхронизации. Например, это имеет место для СНН при использовании фазовой манипуляции: система должна восстановить начальную фазу несущей на фоне весьма значительных изменений того же параметра за счет манипуляции.

Ситуация оказывается особенно сложной, когда применяются современные методы подавления внеполосных излучений (см. гл. 6). Кроме того, в последние годы достигнут значительный прогресс в реализации эффективных систем помехоустойчивого кодирования, так

что нормально радиолиния может работать при отношении сигнал/шум, измеренном в основной полосе сигнала, порядка 1 и менее. Соответственно, при столь значительных шумах должна обеспечиваться и эффективная работа систем синхронизации приемника. Таким образом, функционирование систем синхронизации, использующих в качестве ОС информационный сигнал, существенно усложняется наличием манипуляционных и шумовых помех.

По той же причине в последнее время все чаще применяются варианты, основанные на использовании специальных служебных сигналов. Так, стандарт GSM предусматривает передачу перед началом сеанса связи специального служебного сигнала для обеспечения работы ССН абонентских терминалов. Кроме того, в каждый информационный пакет включается специальное обучающее слово (сообщение), заранее известное на приемном конце и служащее, в частности, для поддержания работы систем синхронизации в течение сеанса.

Что касается структуры систем синхронизации, то различают разомкнутые (рис. 10.1, а) и замкнутые (рис. 10.1, б) системы.

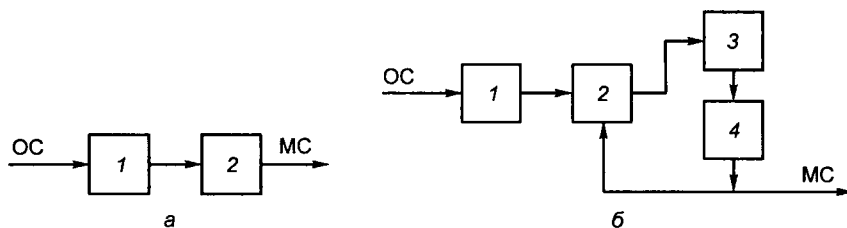


Рис. 10.1. Разомкнутая (а) и замкнутая (б) системы синхронизации

На рис. 10.1, а устройство 1 — измеритель значения подстраиваемого параметра ОС; 2 — управляемый генератор (УГ) местного сигнала (МС), соответствующий параметр которого управляется так, чтобы сделать его по возможности равным измеренному значению. Разомкнутым системам свойственны регулярные ошибки, вызванные непрогнозируемыми изменениями градуировочных кривых измерителей и устройств управления УГ в процессе эксплуатации как за счет старения, так и за счет изменений условий окружающей среды. По этой причине системы синхронизации, использующие только разомкнутые устройства, используются сравнительно редко.

На рис. 10.1, б: 1 — преобразователь ОС; 2 — дискриминатор; 3 — устройство усреднения (УУ) и 4 — УГ. Система работает следующим образом: поступающий ОС предварительно обрабатывается преобразователем, так чтобы по возможности подчеркнуть или выделить содержащиеся в нем данные о регулируемом параметре и ослабить влияние на них флуктуационного шума, всегда сопровождающего ОС, и манипуляции этого сигнала. Конкретные процедуры, выполняемые преобразователем, зависят как от вида ОС, так и от типа рассматриваемой системы синхронизации.

Сигнал с выхода преобразователя вместе с МС поступают на дискриминатор, задача которого — оценка расхождения этих сигналов по анализируемому параметру. Обозначим последний для ОС и МС через x и x^* соответственно. Регулярная часть выходного напряжения дискриминатора — известная функция от ошибки $y = x - x^*$.

На рис. 10.1, б вырабатываемые дискриминатором текущие значения оценки ошибок обрабатываются УУ, в результате чего формируются текущие значения управляющего сигнала, подаваемого на УГ. Для дальнейшего важно различать УГ со статическим и астатическим управлением. В первом случае при снятии управляющего воздействия регулируе-

мый параметр УГ, обозначенный выше как x^* , возвращается к тому значению, которое он имел до подачи воздействия; во втором случае, наоборот — результат воздействия сохраняется после его снятия.

Например, управление частотой автогенератора с помощью подачи напряжения на варикап, подключенный к задающему резонатору, является статическим, поскольку после снятия с варикапа управляющего напряжения его емкость, а вместе с ней и частота генератора возвращаются к исходным значениям. Если же в качестве управляемого генератора использовать цифровой синтезатор, частота сигнала которого, устанавливается путем подачи соответствующей команды, то управление частотой окажется астатическим, поскольку, если новые команды не подаются, то сохраняется значение частоты, установленное последней поданной командой.

МС, формируемый УГ рассматриваемой схемы, с одной стороны, представляет собой выходной сигнал системы синхронизации, а с другой, он подается на дискриминатор, как пояснено выше. В результате и образуется замкнутая схема — система автоматического регулирования (САР), обеспечивающая слежение за интересующим нас параметром x .

10.1.2. Режимы работы САР и их анализ

Анализ САР начинают с отыскания положений равновесия, т.е. таких ее состояний, когда регулярные движения в системе прекращаются. Положение равновесия характеризуются тем, что при обходе по замкнутому контуру регулирования, возникающие напряжения и параметры сигналов оказываются сбалансированными. В зависимости от структуры САР и вида характеристики ее дискриминатора система может иметь ряд положений равновесия. В некоторых из них значение параметра x^* УГ оказывается близким к значению x эталонного сигнала, так что эти положения равновесия соответствуют цели регулирования и являются, поэтому, желательными, а другие — этой цели не соответствуют и являются нежелательными. Какое именно положение равновесия установится в системе и установится ли равновесие вообще, зависит от начальных условий включающих в качестве основного параметра начальную расстройку (значение y в момент замыкания кольца).

Среди положений равновесия могут быть неустойчивые, из которых система уходит под воздействием даже самых малых случайных отклонений. Желательные положения равновесия всегда должны быть устойчивыми.

Работа САР вблизи желательного положения равновесия называется *режимом слежения*. Под действием помех и шумов САР совершает относительно небольшие случайные колебания. Статистические оценки величин соответствующих изменений y определяют точность работы системы синхронизации. В частности, среднее значение этой величины определяет регулярную ошибку регулирования.

Если анализируемый параметр x при работе системы в режиме слежения достаточно медленно изменяется, то система до некоторых пределов отслеживает изменения, не выходя из указанного режима. Когда же пределы, именуемые *полосой удержания*, оказываются превзойденными, слежение прекращается. Полоса удержания — одна из важных характеристик систем синхронизации.

Процесс перехода системы из начального состояния в одно из желательных положение равновесия (точнее в окрестность этого положения равновесия), называется *переходным процессом*. Основной характеристикой переходного процесса следует считать его длительность. Поскольку переходный процесс протекает в присутствии манипуляционных помех и шумов, то его длительность — величина случайная.

Особо следует остановиться на начальной части переходного процесса. Дело в том, что дискриминаторы САР способны вырабатывать сколько-нибудь достоверные данные о значении расстройки u , только если модуль данной величины не превосходит некоторого порогового значения. Только, начиная с таких расстройек, может происходить в дальнейшем целенаправленное движение в системе. На начальном этапе переходного процесса это условие может не выполняться, так что возникает необходимость организации специального процесса поиска области значений параметра, из которой система уже может автоматически двигаться к желательному положению равновесия (режим поиска). Здесь вся область априори возможных значений x заранее разбивается на некоторые интервалы, протяженность которых такова, что если x и x^* попадут в один интервал, то дискриминатор будет выдавать более или менее достоверные данные о расстройке u , так что работа системы автоматически приведет к переходу в режим слежения.

Далее, для организации поиска в состав САР вводится устройство обнаружения режима слежения, призванное фиксировать переход в такой режим. Работа этого устройства основывается на усреднении ошибок, фиксируемых дискриминатором: если модуль ошибки, усредненный за некоторый интервал времени, не превосходит некоторого, принятого за допустимый, предел, то считается, что режим слежения наступил. Сам режим поиска организуется следующим образом: для проверки принадлежности текущего значения x некоторому интервалу значение x^* УГ принудительно выставляется на середину данного интервала. По прошествии времени, необходимого для завершения переходного процесса, с помощью устройства обнаружения слежения определяется правильность выбора данного интервала.

Таким образом перебираются все интервалы и устанавливается тот из них, который приводит к началу слежения. Сам перебор может быть как последовательным, так и параллельным. Последний вариант реализуется, как правило, только в случаях, когда САР использует программируемые вычислительные процедуры на основе микропроцессорных или других современных цифровых устройств. Если синхронизация осуществляется с использованием специальных служебных сигналов, то во избежание чрезмерного увеличения длительности поиска чаще всего также используются параллельные процедуры.

При воздействии на вход САР наряду с опорным сигналом еще и помех, и шумов, в системе, работающей в режиме слежения, время от времени этот режим нарушается, причем система удаляется от желательного положения равновесия и не возвращается в него без реализации специальных процедур (переходит в область притяжения нежелательного положения равновесия или в область параметров, из которой равновесие вообще не достигается). Такое явление именуется *срывом слежения*. Для борьбы со срывами в САР должно применяться устройство обнаружения срывов слежения, которое призвано регистрировать появление срывов и включать процедуры парирования этого явления.

Для фиксации срывов используется то же устройство, что и для обнаружения момента достижения режима слежения. Для этого в соответствующем устройстве вводятся два режима: режим фиксации достижения слежения и режим контроля его наличия. Принцип реализации процедур обнаружения в обоих режимах один и тот же и описан выше. Отличие заключается только в выборе параметров и порогов. Дело в том, что, если система уже находилась достаточно долго в режиме слежения, то вынесение заключения о его исчезновении, требует гораздо большей «осмотрительности», т.е. большего времени анализа, чем первоначальная фиксация этого режима. Следует иметь в виду, что пока САР будет «думать» над тем, не произошел ли срыв слежения, в случаях, когда он действительно произошел, шумы и помехи уведут систему синхронизации весьма далеко от желательного положения равновесия, так что восстановление синхронизма потребует значительного времени и, как правило, повторной реализации длительных процедур поиска.

Нарушение синхронизма в системах синхронизации приемника приводит не только к появлению пачки ошибок соответствующей протяженности в транслируемом сообщении, но, как правило, и к появлению вставок (см. гл. 6) в этом сообщении. Для того, чтобы по возможности исключить такое опасное явление, принимаются специальные меры по парированию срывов слежения. Способность приемника противостоять срывам получила название *обрывоустойчивости*.

В соответствии с изложенным, основной целью мероприятий по обеспечению обрывоустойчивости должно быть недопущение значительных отклонений регулируемых (подстраиваемых) параметров от их требуемых значений при появлении «вспышек» помех и шумов или провалов уровня полезного сигнала. Особое внимание должно уделяться СТС, поскольку именно срывы тактовой синхронизации служат непосредственной причиной возникновения вставок. При обеспечении обрывоустойчивости полезно используется тот факт, что отслеживаемые параметры эталонного сигнала x изменяются весьма медленно по сравнению с длительностью элемента принимаемого радиосигнала. T_3 . Поэтому, если в САР приемника прервать управление параметрами x^* местного генератора (разомкнуть кольцо регулирования) на время nT_3 , то даже для достаточно большого значения n нормальная демодуляция не будет нарушаться.

В связи с этим в устройствах обнаружения срыва вводится режим предупреждения, задачей которого является выдача сигнала предупреждения, как только появляется любое подозрение о возникновении срыва слежения. Основной для данного режима признается быстрота реакции, так чтобы за время, расходуемое на формирование предупреждения, в случае, если оно обосновано, системы синхронизации не успевали существенно изменить параметры x^* . Соответственно, вероятность ложной тревоги при формировании предупреждения может быть весьма высокой, что приводит лишь к тому, что кольцо регулирования изредка размыкается до тех пор, пока устройства обнаружения срывов не проведут подробной проверки факта обрыва. Как правило, обоснованность предупреждения не подтверждается, кольцо снова замыкается и продолжает работу в обычном режиме. Если же факт срыва подтверждается, то работа с разомкнутыми кольцами регулирования сохраняется. При этом синхронизм сохраняется лишь в течение ограниченного времени T_{cc} , определяемого собственными стабильностями частот опорных генераторов приемника и передатчика радиолинии, а также сдвигами частот, вызванными эффектом Доплера.

Если за время T_{cc} обрыв прекратится, и устройство обнаружения слежения это зафиксирует, то кольца регулирования будут замкнуты, и САР перейдет к работе в обычном режиме. Если же это не произойдет, то в момент T_{cc} кольца все равно будут замкнуты, и система начнет работать в том режиме, который используется при первоначальном вхождении в синхронизм, начиная с поиска. В последнем случае неизбежно возникают вставки, и устройства приемника оказываются неспособными их парировать.

Для того чтобы по возможности облегчить необходимый перезапуск внешней по отношению к приемнику аппаратуры, включая оконечную аппаратуру абонента, приемник должен вырабатывать сигналы начала и конца обрыва (в моменты, когда соответствующие факты надежно зарегистрированы).

Остановимся теперь на общем подходе к анализу САР, используемых в системах связи. Дело в том, что работа различных таких систем тесно взаимосвязана. Так обстоит дело, например, применительно к ССН и СТС приемников при фазовой манипуляции. Для обеспечения эффективной работы ССН в ряде случаев должно быть предварительно тем или иным путем обеспечено «снятие манипуляции», т.е. устранение манипуляционных помех. Это возможно только при наличии тактового синхронизма. С другой стороны, пока не обеспечивается синхронизация по несущей, сигнал не может быть качественно протектирован, так

что и нормальная работа СТС не может быть обеспечена. В результате часто приходится фактически объединять эти системы, во всяком случае в процессе поиска. И только после того, как хотя бы грубая синхронизация и по частоте несущей, и по фазе тактовых моментов будет установлена, указанные системы могут начать работать независимо.

Описание САР синхронизирующих даже один параметр сигнала, в условиях, когда остальные уже определены, достаточно сложно; тем более это относится к системам одновременной синхронизации нескольких параметров. В настоящем разделе принят следующий план изложения: вначале рассматривается система синхронизации одного параметра в условиях, когда остальные уже скорректированы и приводятся некоторые основные соотношения, описывающие, прежде всего, режим слежения этих систем. И лишь в заключительной части главы дается качественная характеристика систем совместной синхронизации нескольких параметров. Все описание ведется, прежде всего, применительно к САР с цифровой реализацией, поскольку они занимают господствующее положение в современных системах связи. Кроме того, изложение ориентировано, в первую очередь, на системы с фазовой манипуляцией, включая ее разновидности, описанные в гл. 6, поскольку именно они чаще всего используются в современных системах связи.

10.1.3. Системы фазовой автоподстройки

Система фазовой автоподстройки (ФАП) лежит в основе ряда систем синхронизации, используемых в радиосвязи, таких как ССН, СТС, ССЗ и другие, потому принцип ее действия целесообразно описать специально. Блок-схема системы ФАП представлена на рис. 10.2.

Идеально полезная составляющая ОС представляет собой периодическое колебание, частота которого заранее известна с определенной ошибкой, а фаза — случайна. УГ формирует аналогичное колебание (местный сигнал, МС), частота которого при отсутствии управления отличается от частоты ОС на величину, именуемую *начальной расстройкой* ($F_{\text{н}}$). Дискриминатор схемы вырабатывает сигнал, являющийся известной функцией от разности фаз ОС и МС и потому называется *фазовым дискриминатором* (ФД). Функция, описывающая ФД, определена на интервале от $-\pi$ до $+\pi$. В зависимости от вида этой функции различают релейные (определяющие только знак разности фаз), синусоидальные, треугольные, пилообразные и другие дискриминаторы (рис. 10.3).

Что касается характера управления в кольце ФАП, то здесь всегда предполагается, что в самом УГ осуществляется астатическое управление фазой МС. Обратим внимание на то, что такое управление может рассматриваться и как статическое управление его частотой. Следует, однако, иметь ввиду, что определенные инерционные звенья, включая, например, и идеальные интеграторы, может содержать и УУ, так что в ФАП при необходимости может реализовываться астатическое управление не только фазой МС, но и его частотой.

По принципу действия кольца ФАП, положения его равновесия могут достигаться лишь при равенстве частот ОС и МС. В противном случае разность фаз этих сигналов не может оставаться постоянной, а вместе

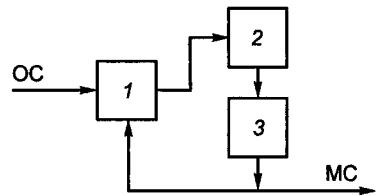


Рис. 10.2. Блок-схема ФАП: 1 — дискриминатор; 2 — УУ; 3 — УГ

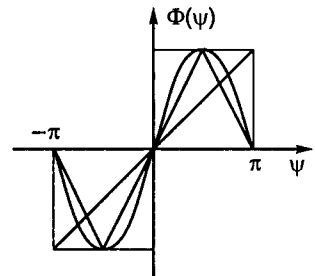


Рис. 10.3. Типовые характеристики ФД

с ней не может оставаться постоянным и сигнал на выходе ФД. Таким образом, остаточная ошибка по частоте в системе ФАП всегда равна нулю. Что касается разности фаз МС и ОС в состоянии равновесия, т.е. остаточной ошибки по фазе φ_0 , то она должна быть такой, чтобы соответствующий сигнал на выходе ФД приводил бы к появлению как раз такого управляющего напряжения на входе УГ, которое бы обеспечивало компенсацию начальной расстройки F_n . Это означает, в частности, что если в состав УУ входит идеальный интегратор, то и остаточная ошибка по фазе будет нулевой.

Изложенное позволяет утверждать, что в системе ФАП φ_0 тем больше, чем больше F_n . Исключения составляют лишь случаи использования в УУ идеальных интеграторов, когда $\varphi_0 = 0$. Практически для всех рассматриваемых систем синхронизации должно выполняться условие $|\varphi_0| \ll 1$. Поэтому всегда, когда можно ожидать достаточно больших начальных расстроек по частоте, приходится использовать в УУ идеальный интегратор. Следует специально остановиться на вопросе о том, что в данном контексте имеется в виду под «достаточно большим». Если приводить качественные оценки, то речь здесь идет о следующем: пусть мы имеем кольцо с длительностью переходного процесса $T_{\text{пн}}$. Для такого кольца начальная расстройка по частоте такая, что $F_n T_{\text{пн}} \ll 1$ является малой; в противном случае расстройка должна быть признана большой.

Случай малых F_n представляется типовым для СТС. Наоборот, для ССН типовым оказывается случай больших F_n .

10.2. Системы синхронизации по несущей и тактам при фазовой манипуляции

10.2.1. Цифровые представления входных данных систем синхронизации

Как уже отмечалось, рассматриваем только цифровые варианты реализации САР. Таким устройствам всегда предшествуют АЦП, которые преобразуют входные радиосигналы в цифровую форму. Возможные варианты реализации АЦП радиосигналов проанализированы в 1.3. Там же были исследованы и варианты, позволяющие получить цифровые представления амплитуды и фазы радиосигналов $A(t)$ и $\varphi(t)$. Ниже будем полагать, что такие представления ОС на входе САР получены, так что при цифровой обработке радиосигналов используются наборы чисел $A_k = A(t_0 + kT_d)$ и $\varphi_k = \varphi(t_0 + kT_d)$ для любых целых k положительных и отрицательных. Здесь t_0 — начальный сдвиг моментов отсчета по времени, а T_d — шаг дискретизации по времени.

В общем случае T_d не связано непосредственно с первоначальной частотой дискретизации F_0 : цифровое представление сигнала позволяет вычислить его значения в любые моменты времени и, в частности, получить значения A_k и φ_k для любых T_d и t_0 .

Интервал T_d выбирают так, чтобы в длительности элементарной посылки укладывалось целое число таких интервалов. Это число бывает небольшим: используют всего несколько замеров амплитуды и фазы за каждую посылку. Для того, чтобы такие отсчеты могли быть эффективно реализованы при построении демодуляторов и САР приемников, величина t_0 должна выбираться так, чтобы момент взятия хотя бы одного отсчета за посылку соответствовал бы истинным тактовым точкам принимаемого сигнала. Таким образом, величина t_0 должна определяться работой системы тактовой синхронизации (СТС) приемника.

10.2.2. Тактовая синхронизация

Рассмотрение систем синхронизации по несущей и тактам в пределах общей главы определяется следующими обстоятельствами: *во-первых*, работа этих систем основана на анализе одного и того же ряда данных, прежде всего отсчетов φ_k . *Во-вторых*, как уже отмечалось выше, работа этих систем взаимосвязана; в общем случае СТС не может решить свою задачу, пока ССН не достигла режима слежения, и наоборот. В связи с этим режим поиска для обеих систем должен быть совместным. Наконец, обе системы синхронизации представляют собой различные реализации одной и той же САР — системы ФАП.

Рассмотрим построение СТС и ССН применительно к приему фазоманипулированных m -ичных сигналов, причем номинальные значения фаз элементарных сигналов расположены равномерно на интервале 2π с шагом $q = 2\pi/m$. Для приема таких сигналов используются системы синхронизации различной структуры. В задачи настоящего изложения не входит сколько-нибудь полная их классификация и описание. Ниже будет кратко представлен один типовой вариант построения ССН и СТС.

Начнем с описания принципов извлечения информации о положении тактовых точек из манипулированного радиосигнала.

Обозначим длительность элементарного сигнала через T_3 и примем для определенности изложения, что по номиналу $F_0 = 2/T_3$, т.е. частота дискретизации выбрана так, чтобы получать два отсчета амплитуды и фазы сигнала за посылку. Отметим, что такой выбор частоты дискретизации требует, чтобы полоса частот вне которой спектр сигнала можно считать равным нулю, удовлетворяла условию $\Delta f \leq 2B$, где скорость манипуляции $B = 1/T_3$.

Важно подчеркнуть, что, поскольку моменты дискретизации в приемнике определяются местной тактовой последовательностью приемника (МТП ПРМ), а временные положения истинных тактовых точек принимаемого сигнала — аналогичной последовательностью передатчика (МТП ПРД), то при разомкнутом кольце СТС всегда наблюдается большее или меньшее расхождение частот этих последовательностей, так что моменты взятия отсчетов постоянно сдвигаются относительно посылки принимаемого сигнала. Задача СТС и заключается в том, чтобы устранить это перемещение путем подстройки частоты МТП ПРМ под тактовую частоту принимаемого сигнала. Кроме того, фаза МТП ПРМ должна быть установлена так, чтобы совместить моменты появления, например, четных отсчетов с началом посылок, а нечетных — с их серединами.

Рассмотрим истинные тактовые точки принимаемого полезного ФМ-сигнала. Те из них, в которых фактически передаваемые элементарные посылки не изменяются, не дают информации для работы СТС и, потому названы пассивными. Наоборот, те тактовые точки, где фактически происходит изменение передаваемой посылки, именуется *активными*. При прохождении ФМ-сигнала через любой полосовой фильтр появление активной тактовой точки всегда приводит к падению амплитуды выходного сигнала фильтра, вызванному затуханием отклика фильтра на предыдущую посылку, при еще не установившемся отклике на последующую. В частности, если в этой точке фаза радиосигнала за счет манипуляции изменяется на 180° , то амплитуда сигнала на выходе фильтра проходит через ноль (см. 6.2). Отсюда следует также, что именно в точках с минимальной амплитудой выходного сигнала фильтра наиболее круто изменяется его фаза. Поэтому именно такие моменты времени и должны рассматриваться как истинные тактовые точки сигнала на выходе фильтра. Наиболее надежные указания на временное положение тактовых точек могут быть получены путем анализа закона изменения фазы сигнала на выходе фильтра, ибо провалы его амплитуды могут возникать и по причинам, не связанным со сменой передаваемого элементарного сигнала.

Дискретизация усложняет извлечение из эталонного сигнала информации о временных положениях тактовых точек: в случаях, когда берется всего несколько отсчетов за посылку трудно непосредственно судить, например, о местоположениях моментов наиболее крутого изменения фазы и т.п. В рассматриваемом случае двух отсчетов за посылку, обработка данных о фазе φ_k для управления работой СТС сводится к следующему: последовательность точек отсчета рассматривается как чередование главных и вспомогательных точек. Например, в качестве главных выбирают отсчеты с четными номерами $k = 2j$. Далее, считается, что если модуль разности фаз между двумя последовательно расположенными главными тактовыми точками $\alpha_j = |\varphi_{2j} - \varphi_{2(j-1)}|$ соизмерим с q , например, превосходит $q/2$, то на j -м тактовом интервале имела место активная тактовая точка и относящиеся к данному интервалу данные должны использоваться СТС; в противном случае интервал считается пассивным и соответствующее значение α_j СТС не учитывается.

Учет значений разностей фаз на интервалах, признанных активными, осуществляется следующим образом: рассматривается значение фазы сигнала во вспомогательной точке отсчета этого интервала $k = 2j - 1$, и оно сравнивается с ее значениями в соответствующих главных точках: j -й и $(j - 1)$ -й. Если более близким оказывается значение фазы в j -й точке, то вырабатывается сигнал минус 1, а в противном случае — плюс 1. В идеале, когда главные отсчеты соответствуют моментам завершения переходного процесса в фильтре, вызванного сменой передаваемой посылки, фазы сигнала во вспомогательных точках отсчета должны лежать ровно посередине между фазами в ближайших к ним главным точкам отсчета. Поэтому появление сигнала плюс 1 означает необходимость сдвига МТП ПРМ в направлении к j -му главному отсчету. В противном случае сдвиг должен осуществляться в противоположном направлении. Приведенный принцип выделения информации об ошибке тактовой синхронизации одновременно служит описанием принципа действия релейного фазового дискриминатора для СТС. Именно такой дискриминатор и используется в рассматриваемой системе.

Сама СТС при таком дискриминаторе функционирует следующим образом: выбирается некий достаточно малый шаг коррекции фазы $\delta \ll \pi$ и на каждом признанном активным тактовом интервале фаза МТП ПРМ корректируется на эту величину в ту или иную сторону в соответствии с сигналом, поступающим с выхода дискриминатора. Значение шага коррекции определяет инерционность СТС, чем меньше δ , тем больше инерционность и тем меньше влияют на работу системы шумы и помехи. Типовое значение шага коррекции имеет порядок $0,01\pi$ рад. Соответственно, длительность переходного процесса для такой СТС составляет в среднем несколько сот тактов.

Изучаемая СТС реализует простейшую систему ФАП, в которой применен релейный дискриминатор и вообще не используется какое-либо специальное УУ. В такой системе ФАП, как отмечалось выше, должна существовать в установившемся режиме остаточная ошибки по фазе, пропорциональной начальной расстройке по частоте. Однако, как правило, для СТС начальная расстройка по частоте следования МТП ПРМ весьма мала. Действительно, с учетом всех дестабилизирующих факторов, включая сдвиги частоты, вызванные эффектом Доплера, относительная начальная расстройка по частоте для СТС не превосходит 10^{-4} . Вызванный такой расстройкой набег фазы за тактовый интервал не более $2\pi 10^{-4}$ рад, что во много раз меньше приведенного выше типового значения шага коррекции. Это означает, что указанная начальная частотная расстройка не влияет практически на работу СТС. Именно поэтому в качестве такой системы и может быть использованная простейшая схема ФАП.

10.2.3. Синхронизация по несущей

Перейдем теперь к описанию работы ССН. Информация для работы такой системы извлекается из значений фазы φ_j в главных точках отсчета. Дополнительно ССН может использовать и данные A_j .

Прежде чем отсчеты фазы будут использованы для работы ССН, они должны быть подвергнуты специальному преобразованию, называемому *снятием манипуляции*. Дело в том, что значения φ_j , кроме интересующей нас информации о несущей частоте и начальной фазе принимаемого сигнала, зависят также и от значения манипулирующего параметра на данном тактовом интервале. Последняя зависимость служит помехой в работе ССН и ее желательно, по возможности, устранить. Эта процедура выполняется по следующему алгоритму.

Разделим интервал всех возможных значения φ_j шириной в 2π на Q равных секторов. Пусть первый из них занимает углы от $-q/2$ до $q/2$ (в качестве области определения фазы примем интервал $-\pi \dots +\pi$). Всякое попадание φ_j в некоторый сектор можно рассматривать как следствие наложения манипуляционного сдвига на значение фазы сигнала, лежащей, например, в первом секторе. (Какой именно сектор принимается здесь за первый, значения не имеет, поскольку по самому манипулированному сигналу, фаза его несущей может быть восстановлена принципиально только с точностью до целого числа интервалов q). Поэтому процедура снятия манипуляции должна сводиться к вычитанию или прибавлению к значению φ_j целого числа интервалов q , причем это число должно быть выбрано так, чтобы результат ψ_j принадлежал первому сектору. Именно величины ψ_j и должны использоваться ССН для оценки несущей частоты и начальной фазы принимаемого радиосигнала.

Ниже будет описана САР, выполняемая в соответствии с принципами работы системы ФАП, но чисто вычислительным путем. Принимается, что на вход соответствующего вычислителя поступают отсчеты фазы ОС ψ_j , на основе которых вычислитель вырабатывает текущие поправки по частоте и фазе. Будем считать, что отсчеты фазы ψ_j осуществляются по отношению к сформированному в приемнике из сигнала его опорного генератора исходному гармоническому колебанию (ИК), частота которого f_n совмещена по номиналу с несущей ОС и в процессе работы ССН не меняется. Поскольку речь идет о цифровом вычислителе, то на выходе он выдает цифровые отсчеты поправок по частоте и фазе (по отношению к параметрам ИК) в такте поступления отсчетов фазы ψ_j . Обозначим их через θ_j (поправка по фазе) и f_j (поправка по частоте) и будем именовать их *текущей фазой* и *частотой* соответственно. Заметим, что данные величины содержат регулярные и случайные составляющие. Регулярная часть этих величин и определяет полезную компоненту выработанной САР местной несущей частоты, случайные — отображают погрешности работы системы, вызванные воздействием на нее внешних помех и шумов.

Для пояснения соответствия текущих частоты и фазы параметрам несущей приведем некоторые примеры. Не подстроенной местной несущей — ИК с частотой f_n и фазой нуль соответствуют нулевые текущая частота и текущая фаза: $f_j = 0$; $\theta_j = 0$. Истинная несущая ОС имеет частоту $f_n = f_n + F_n$, где F_n — начальная расстройка. Поэтому истинной несущей соответствует текущая частота F_n . А каково значение текущей фазы, соответствующей истинной несущей? Напомним, что, как отмечено выше, текущие параметры измеряются относительно ОС, так что для истинной несущей $\theta_j = 2\pi F_n t_j + \varphi_n$, где φ_n — начальная фаза истинной несущей, измеренная относительно частоты f_n . Таким образом, текущая фаза истинной несущей постоянно изменяется ($\theta_{j+1} - \theta_j = 2\pi F_n T_d$), что определяется просто расхождением средних частот ОС и ИК.

В соответствии со структурой ФАП, фазовый детектор должен вырабатывать на каждом такте определенную функцию от разности очередного отсчета фазы и текущей фазы:

$z_j = \Phi(\psi_j - \theta_j)$. Отсчеты ошибок z_j в системе ФАП поступают на УУ, которое вырабатывает сигналы, корректирующие в общем случае и частоту и фазу местной несущей. Будем полагать, что алгоритм работы УУ сводится к следующему: пусть по очередному поступившему на его вход значению ошибки z_j УУ выдает команду на коррекцию фазы местной несущей на величину $k_\Phi z_j$, где $k_\Phi < 1$ — некоторый коэффициент. Кроме того, УУ формирует сигнал оценки ошибки по частоте путем суммирования ошибок z_j с некоторым коэффициентом $k_f \ll 1$. Такая оценка ошибки по частоте исходит из следующих соображений: если бы ошибка по частоте отсутствовала, то ошибки по фазе z_j после завершения переходного процесса определялась бы только шумами и помехами, и равновероятно принимала бы положительные и отрицательные значения, так что в среднем суммирование этих ошибок давало бы нулевой результат. Если же имеет место ошибка по частоте, то в величину ошибки z_j входит регулярная составляющая, определяемая набегом фазы из-за расхождения частот. Суммирование ошибок z_j и должно выявить эту составляющую.

В результате уравнение, описывающее работу ФАП, принимают вид

$$\theta_{j+1} = \theta_j - k_\Phi \Phi(\psi_j - \theta_j) + 2\pi T_d f_j, \text{ где } f_{j+1} = f_j + k_f \Phi(\psi_j - \theta_j). \quad (10.1)$$

Это соотношение показывает, что изменение текущей фазы на очередном такте работы кольца характеризуется как величиной, измеренной на этом такте фазовой ошибки (отклонением измеренного на данном такте значения фазы ОС от текущей фазы), так и коррекцией текущей фазы в соответствии с установленным к данному такту значением текущей частоты. Что касается самой текущей частоты, то она также корректируется на каждом такте работы кольца на величину, определяемую измеренной фазовой ошибкой.

Если второе соотношение (10.1) подставить в первое, то убедимся, что речь идет о разностном уравнении второго порядка. В частном случае можно оценку частотной погрешности не использовать, принимая $f_j = 0$. Тогда мы придем к разностному уравнению первого порядка. Такие системы используют при сравнительно небольших начальных расстройках; в общем случае применяется система с оценкой и по частоте.

Вычисление основных параметров ССН, описываемой уравнением (10.1), представляет собой достаточно сложную задачу, выходящую за рамки настоящего изложения. Отметим лишь те из них, которые могут быть получены элементарными средствами.

Начнем с системы первого порядка. Ее положение равновесия соответствует сохранению постоянного, в среднем, напряжения на выходе фазового детектора, т.е. выражается соотношением

$$\overline{\theta_{j+1} - \theta_j} = k_\Phi \overline{\Phi(\psi_j - \theta_j)} = \text{const}. \quad (10.2)$$

Для того чтобы оно выполнялось, необходимо и достаточно, чтобы $\psi_j - \theta_j = \text{const}$. Поскольку фаза ψ_j представляет собой замеры фазы ОС с несущей частотой f_n относительно ИК с частотой f_n , то можно утверждать, что $\overline{\psi_{j+1} - \psi_j} = 2\pi F_n T_d$. Поэтому выполнение (10.2) означает, что ту же компоненту содержит и θ , так что

$$\overline{\theta_{j+1} - \theta_j} = 2\pi F_n T_d. \quad (10.2a)$$

Таким образом, в положении равновесия частоты местной несущей и ОС совпадают, а разности фаз этих колебаний выражаются в виде

$$\overline{\Phi(\psi_j - \theta_j)} = \frac{2}{k_\Phi} \pi F_n T_d. \quad (10.3)$$

Это уравнение имеет два решения, поскольку рассмотренные выше (см. рис. 10.3) характеристики ФД пересекаются горизонтальными прямыми в двух точках, одна из которых

лежит на нисходящей, а другая — на восходящей ветвях данной характеристики. Одно из двух положений равновесия оказывается устойчивым и, следовательно, рабочим, а другое неустойчивым и нерабочим. Аргумент характеристики ФД, удовлетворяющий этому соотношению, определяет остаточную ошибку по фазе в ССН.

Выражение (10.3) позволяет утверждать, что полоса удержания для системы первого порядка характеризуется соотношением

$$F_n \leq \frac{k_\phi}{2\pi T_d}. \quad (10.4)$$

Когда начальная расстройка гораздо меньше границы, определяемой этим соотношением, ошибка по фазе оказывается достаточно малой, так что использование системы ФАП первого порядка можно считать оправданным. Следует иметь в виду, что задача борьбы с помехами требует, как уже отмечалось выше, чтобы коэффициент $k_\phi \ll 1$. Это требование и ограничивает достижимую в системе первого порядка полосу удержания F_e величиной порядка $10^{-3}F_s$, где F_s — частота следования элементарных посылок в эталонном сигнале (в рассматриваемом случае $F_s = 1/T_d$).

Что касается полосы захвата F_p , то для системы первого порядка она близка к полосе удержания.

Если $F_n \ll F_p$, то оправдано применение системы первого порядка. В противном случае необходимо применять систему второго порядка.

В такой системе положения равновесия определяются выражениями:

$$\overline{f_j} = F_n; \quad \overline{\Phi(\psi_j - \theta_j)} = 0, \quad (10.5)$$

так что выполняется и (10.2а). Эти соотношения показывают, что в системе второго порядка состояние равновесия соответствует равенству частот местной несущей и эталонного сигнала, а среднее значение ошибки по фазе $\overline{\psi_j - \theta_j} = 0$.

Полоса удержания в системе второго порядка теоретически не ограничена. Что же касается полосы захвата, то она имеет тот же порядок, что и в системе первого порядка с тем же коэффициентом k_ϕ . Поэтому здесь, как правило, необходимо использовать поиск по частоте.

10.2.4. Организация процессов поиска

Как уже отмечалось, поиск может осуществляться как за счет периодического включения на передающей стороне радиолинии в манипулирующий сигнал специального контрольного (обучающего) слова, так и без передачи специальной служебной информации. Возможен вариант, когда ОС передается параллельно основному сигналу. Во всех случаях поиск может быть параллельным или последовательным. При параллельном поиске в память записывается необходимый отрезок ОС и в сопоставлении с ним проверяется весь набор гипотез о значении анализируемого параметра. Предполагается, что быстрдействие соответствующего анализатора таково, что проверка всех гипотез занимает время, соизмеримое с длительностью записанного отрезка ОС. В случае, если это условие не выполняется, запоминание отрезка ОС теряет смысл, поскольку за время проверки одной гипотезы успевает поступить очередной образец ОС. Когда для проверки каждой очередной гипотезы используется новый такой образец, говорят о последовательном поиске. Параллельный поиск используется, чаще всего, при работе по обучающему слову.

Применительно к ССН и СТС в последовательном варианте ограничиваются только поиском по частоте. Поиск по такту и фазе несущей вообще не осуществляется. Время про-

верки каждой гипотезы о частоте несущей выбирается достаточно большим, так чтобы в случае, когда гипотеза оказывается правильной, успели завершиться переходные процессы как по такту, так и по фазе несущей. Сам поиск по частоте организуется следующим образом: весь интервал значений возможных начальных расстройок от $-F_0 \leq F_n \leq F_0$ (рис. 10.4) разбивается на подинтервалы шириной по $F_{ш}$ каждый ($F_{ш} \leq F_0$). Смысл дальнейших процедур заключается в проверке гипотез о принадлежности истинного значения начальной расстройки тому или иному подинтервалу. Последние нумеруются, как показано на рис. 10.4, когда четные номера имеют интервалы одной половины общего интервала, а нечетные — другой («веерная» нумерация). Такая нумерация, если ей следовать при определении очередности проверки гипотез, сокращает среднее время поиска, поскольку, как правило, априорная вероятность пребывания истинного значения начальной расстройки в том или ином подинтервале убывает по мере его удаления от нуля.

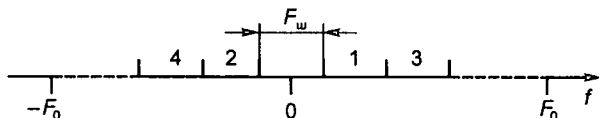


Рис. 10.4. Организация поиска по частоте в ССН

Сама проверка гипотезы по поводу j -го интервала осуществляется следующим образом: при разомкнутом кольце регулирования собственная частота местной несущей принудительно сдвигается на величину минус $0,5j$ (для четных j) или на плюс $0,5(j+1)$ для нечетных. Далее кольцо замыкается на определенное время, необходимое для завершения переходного процесса, в случае если захват действительно происходит. По истечении этого времени оценивается наличие синхронизма. Если он не обнаруживается, то проверяемая гипотеза считается не подтвердившейся, и система переходит к проверке гипотезы относительно $(j+1)$ интервала.

Если фиксируется наличие синхронизма, то система переходит в режим слежения, при котором постоянно проводятся проверки наличия синхронизма. В случае, если несколько проверок подряд укажут на отсутствие синхронизма, то поиск начинается вновь с того значения j , для которого был обнаружен захват.

Алгоритм поиска опирается на возможность обнаружения факта наличия или отсутствия синхронизма. Как эти факты могут быть установлены? Возможны различные способы решения данной задачи. Целесообразно выбрать среди них такой, для которого время анализа было бы не слишком большим, поскольку, как это ясно из вышеизложенного, в процессе поиска такой анализ приходится повторять многократно. Чаще всего на практике используется вариант обнаружения синхронизма, основанный на том, что в синхронном режиме в среднем отклонения замеров фазы (со снятой манипуляцией) ψ_j от текущего значения ее оценки θ_j , выработанной системой ФАП, должны быть меньшими этих отклонений при отсутствии синхронизма. В соответствии с этим вариантом в ССН формируется величина

$$G = \sum_j |\theta_j - \psi_j|, \quad (10.6)$$

где суммирование распространяется на все отсчеты, принадлежащие интервалу оценки наличия синхронизма. Далее значение G сравнивается с некоторым, надлежащим образом выбранным, порогом и, если она оказывается ниже порога, то принимается решение о наличии синхронизма, в противном случае — о его отсутствии. Поскольку время анализа

приходится жестко ограничивать, имеется достаточно высокая вероятность ошибочного заключения о наличии или отсутствии синхронизма. Поэтому перед тем как окончательно принять решение о прекращении поиска время анализа специально увеличивается в несколько раз, а в случае, когда полный цикл поиска (по всем j) не обнаруживает синхронизма, эти циклы повторяются вновь и вновь.

Остановимся на алгоритме поиска по обучающему слову. Отметим, прежде всего, что обучающее слово должно выбираться таким образом, чтобы позволять как можно точнее определять свое временное положение. Говоря качественно, необходимо потребовать, чтобы оно было как можно меньше похоже на свои собственные сдвиги во времени, что сделает менее вероятными вызванные воздействием шумов ошибки, при которых эти сдвинутые положения принимаются за истинные.

Степень «похожести» некоторой последовательности символов $d(j)$, заданной для j от 1 до N на свои сдвиги определяется корреляционной функцией

$$R(k) = \sum_j d(j)d(j-k), \quad (10.7)$$

причем в этой сумме члены с номерами j , для которых один из сомножителей не определен не учитываются.

Например, если в качестве обучающего выбрать слово, состоящее из повторений одного и того же символа, то ошибки оценки его временного положения на несколько символов окажутся весьма вероятными, поскольку при $k \ll N$ корреляционная функция $R(k)$ близка к $R(0) = N$. Аналогичный результат получится, если в качестве обучающего слова выбрать «меандр» (периодическое чередование 1 и -1); в этом случае высоковероятными будут ошибки на некоторое четное число символов, поскольку и здесь при $k \ll N$ корреляционная функция $R(2k)$ близка к $R(0) = N$. С другой стороны, можно привести примеры комбинаций, резко меняющихся при сдвиге. Таковыми являются, в частности, рассмотренные в гл. 7 последовательности Баркера. Напомним, что при сдвиге на единицу и более символов для этих последовательностей $R(k)$ не превосходит по модулю 1. Однако для $N > 13$ таких последовательностей не найдено. Если же не ограничивать столь жестко значений корреляционной функции при $k \neq 1$, потребовав лишь, чтобы по модулю они не превосходили \sqrt{N} , то многие двоичные последовательности окажутся «хорошими». Действительно, в соответствии с центральной предельной теоремой теории вероятностей, для достаточно большого N выбранная наугад последовательность соответствует указанному требованию с вероятностью близкой к единице.

Пусть теперь обучающая последовательность выбрана и заранее известна на приемном конце радиолинии. Поскольку известен также и используемый в радиолинии способ манипуляции, то оказывается известным и сигнал, соответствующий данной последовательности. Однако следует обратить внимание на то обстоятельство, что имеющиеся на приеме сведения об этом сигнале практически всегда являются неполными. Так, неизвестными являются временные положения начала этого сигнала и его тактовых точек на входе приемника, а также точное значение его несущей частоты и ее фазы относительно местного генератора приемника. Этот факт определяется неизбежным расхождением часов и нестабильностями опорных генераторов приема и передачи, а также обычно неконтролируемым запаздыванием сигнала в радиолинии, вызванным конечной скоростью распространения радиоволн. Перечисленные параметры сигнала и должны быть определены в первом приближении по сигналу, отображающему обучающее слово. Именно в этом и заключается задача поиска по такту и несущей с использованием обучающего слова.

Опишем в общих чертах рациональный путь решения этой задачи. Итак, пусть задан соответствующий обучающему слову сигнал (обучающий радиосигнал ОбРС), $S(t_0, f_n, \theta, t)$,

где t — время, а t_0 , f_n и θ (момент начала сигнала, его несущая частота и ее начальная фаза соответственно) — параметры сигнала, относительно которых известны лишь их априорные распределения вероятностей. Указанный набор параметров сигнала соответствует принимаемому здесь предположению, что канал связи является стационарным, так что форма сигнала на выходе канала может быть определена заранее. Для каналов с многолучевостью принимаемый сигнал зависит еще и от характеристик канала, носящих случайный характер. В последнем случае параметры канала также должны определяться по обучающему слову; соответствующее устройство, называемое эквалайзером (см. гл. 9).

Близким к идеальному алгоритмом оценки параметров ОБРС является, как это следует из теории потенциальной помехоустойчивости, алгоритм максимального правдоподобия. В соответствии с данным алгоритмом в качестве оценок параметров должны быть выбраны их значения, максимизирующие интеграл

$$I = \int_{t_0}^{t_0+T_{OC}} z(t)S(t_0, f_n, \theta, t)dt, \quad (10.8)$$

где $z(t)$ — напряжение на входе приемника, представляющее собой сумму переданного ОБРС и флуктуационного шума канала связи; T_{OC} — длительность ОБРС.

На практике выполнить строго процедуры отыскания максимума выражения (10.8) невозможно, поскольку каждый параметр может принимать любые значения из соответствующего интервала, то пришлось бы сопоставлять бесконечное количество значений I . Поэтому реально приходится в области значений каждого параметра выбирать некоторую дискретную сетку его значений и вычислять значения I , при всевозможных сочетаниях параметров, принимающих значения на таких сетках. Сопоставляя получающиеся значения I и находят то сочетание параметров, для которого I близко к максимальному значению.

Ясно, что чем меньше шаги указанных сеток, тем точнее будут получаемые таким путем оценки значений параметров. С другой стороны, сокращение ширины шагов сеток приводит к быстрому возрастанию объема вычисления, которые должны быть выполнены. Поэтому шаги сеток должны выбираться оптимальными, так чтобы не снижать существенно точности оценок и не приводить к неоправданно большим объемам вычислений. При этом требуемая точность оценки параметров должна определяться исходя из данных о влиянии величин ошибок синхронизации на качество приема информационных сигналов.

Ниже описана одна из возможных реализаций устройства поиска по ОБРС и приведены ориентировочные значения ее характеристик. Устройство содержит два регистра сдвига, каждый элемент памяти которых хранит один отсчет сигнала. Оказывается достаточным представлять каждый такой отсчет числом, содержащим три-четыре двоичных разряда. Такты записи и считывания, подаваемые на регистры, следуют с интервалом взятия отсчетов входного сигнала, который в два раза меньше длительности элементарной посылки. В первый регистр записываются числа $c_j = A(t_j)\cos\varphi(t_j)$, а во второй — числа $s_j = A(t_j)\sin\varphi(t_j)$. Здесь $A(t_j)$ и $\varphi(t_j)$ — последовательность отсчетов соответствующих параметров входного напряжения приемника (с учетом шумов).

Число элементов памяти каждого из регистров m в два раза превосходит число элементарных посылок, образующих ОБРС. При этом в регистре на j -м такте хранятся числа с номерами от $j - m + 1$ до j . Выходы всех элементов памяти регистров подаются на вычислитель, в памяти которого хранятся цифровые представления ОБРС для следующих значений параметров: $t_0 = 0$, $f_n = k\delta f$; $\theta = \pi/2$ и $\theta = 0$. Шаг сетки по частоте выбирается порядка $\delta f = 1/(8T_{OC})$. Интервал значений k определяются максимально возможной начальной расстройкой по частоте F_0 : $|k| \leq n = F_0/(\delta f)$, причем в этом выражении значение последней

дроби округляется до ближайшего меньшего целого числа. При таких значениях параметров, хранимые в вычислителе цифровые представления ОБРС, имеют вид:

$$C_{k,j} = A(t_j) \cos[2\pi k \delta f t_j + \varphi(t_j)] \text{ для } \theta = 0; \quad (10.9)$$

$$S_{k,j} = A(t_j) \sin[2\pi k \delta f t_j + \varphi(t_j)] \text{ для } \theta = \pi/2. \quad (10.10)$$

Здесь j принимает значения от 1 до m .

Далее вычислитель для каждого j вычисляет наборы чисел:

$$a_{k,i} = \sum_{i=1}^m c_{j-m+i} C_{ki} \text{ и } b_{k,j} = \sum_{i=1}^m s_{j-m+i} S_{ki}. \quad (10.11)$$

Наконец, коэффициенты правдоподобия определяются так:

$$\lambda_{k,j} = \sqrt{a_{k,j}^2 + b_{k,j}^2}. \quad (10.12)$$

На каждом такте работы схемы (для каждого j) из всех найденных на этом такте значения коэффициентов правдоподобия выбирается максимальный (по k). Пусть он имеет значение λ_j и номер $k = l$. Значение λ_j сравнивается с некоторым заранее выбранным порогом: если оно превышает этот порог, то считается, что ОБРС обнаружен и в качестве параметров ОБРС принимаются значения:

$$t_0 = t_j; F_n = l \delta f \text{ и } \theta = \arctg(b_{l,j} / a_{l,j}) \text{ при } b_{l,j} > 0, \text{ или } \theta = \arctg(b_{l,j} / a_{l,j}) + \pi \text{ при } b_{l,j} < 0.$$

Если λ_j не превышает порога обнаружения, то считается, что ОБРС на данном такте не обнаружен.

Поскольку вставки обучающих последовательностей в информационный поток осуществляются периодически, то если данный ОБРС при описанном алгоритме будет, из-за воздействия шумов, пропущен, то будет обнаружен один из следующих ОБРС. Однако если число символов в обучающей последовательности составляет несколько десятков и порог обнаружения выбран рационально, то с высокой вероятностью ОБРС будет обнаружен при первом же его прохождении через описанное устройство. Определенные оценки параметров сигнала позволяют СТС и ССН приемника сразу же перейти в режим слежения.

10.2.5. Системы поиска и синхронизации по задержке для широкополосного сигнала

К устройствам поиска, использующим ОБРС, относится и система поиска по задержке, входящая в состав приемников широкополосного сигнала (ШПС). Задача, решаемая этой системой, была описана в 6.2. В качестве обучающего сигнала здесь выступает та ПСП, которая была использована при формировании ШПС и переносится этим сигналом постоянно, параллельно с информацией. В связи с такой непрерывной передачей обучающего сигнала здесь используется последовательный поиск. Напомним, что ПСП представляет собой двичную цифровую периодическую последовательность, вид которой заранее известен на приемной стороне радиолинии и может быть воспроизведен в приемнике в виде местной ПСП $v(t)$. Для обеспечения свертки ШПС по частоте необходимо обеспечить синхронизм этой ПСП с ПСП принимаемого сигнала $v_c(t)$ как по такту, так и по циклу. Если запаздывание t_s между $v(t)$ и $v_c(t)$ по модулю приближается к τ_0 (τ_0 — длительность бита ПСП) или превосходит эту величину, то свертка ШПС не происходит, причем отсутствует какая-либо возможность получить сведения о величине данного запаздывания. Соответствующий фазовый дискриминатор может функционировать только при $|t_s| < \tau_0$, когда уровень сигнала свертки возрастает по мере уменьшения величины запаздывания, что и может служить индикатором величины ошибки регулирования.

В то же время, начальная расстройка по задержке может составлять величину порядка цикла ПСП, который во много раз превосходит τ_0 . Таким образом, тут необходим поиск по задержке. Используется поиск с шагом $\tau_0/2$, осуществляемый путем скачкообразного изменения на эту величину такта местной ПСП. Сдвиг осуществляется все время в одну и ту же сторону, так что одновременно с тем же шагом осуществляется и поиск по циклу. После каждого шага поиска фаза местной ПСП не изменяется в течение некоторого времени (назовем его *временем анализа*), необходимого для достаточно надежной регистрации результата последнего шага поиска. Искомым результатом должно явиться появление свертки сигнала по частоте. Для регистрации этого факта используется обнаружитель, состоящий из полосового фильтра (ПФ), с полосой пропускания, равной полосе полезного сигнала после его свертки (полосе узкополосного манипулированного сигнала), амплитудного детектора (АД) и порогового устройства (ПУ). На рис. 10.5 показаны три таких включенных параллельно тракта. Верхний из них после достижения синхронизма по задержке используется для приема информационного сигнала, а до этого момента вместе с двумя другими вспомогательными трактами — задействован в процессе поиска и синхронизации. Входящие в состав представленных трактов ПУ, должны иметь уровни срабатывания, соответствующие появлению свертки по частоте на выходе ПФ. В основном тракте с выхода ПФ сигнал с одной стороны поступает на демодулятор, а с другой на АД с ПУ; это обеспечивает возможность работы тракта в режиме поиска по задержке.

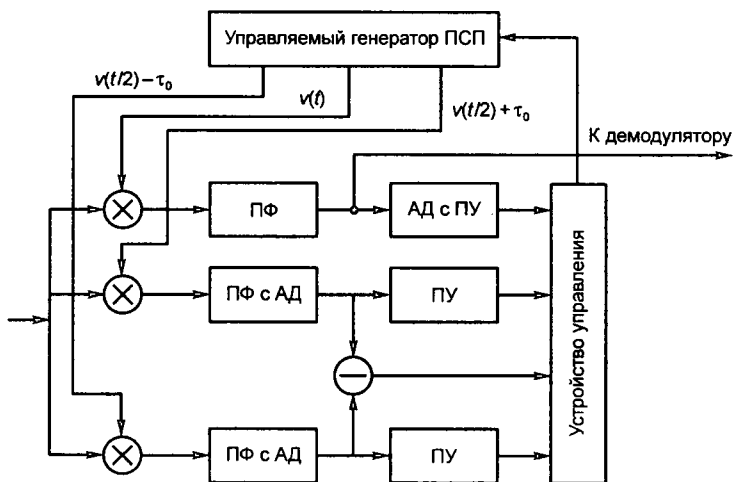


Рис. 10.5. Блок-схема системы поиска и синхронизации по задержке

Следует обратить внимание на фазы ПСП, поступающих на каждый из трактов: фазы ПСП, подаваемых на вспомогательные тракты отличаются на такт, а ПСП основного тракта — лежит между ними.

Указанные три тракта, работая в режиме поиска в параллель, на каждом шаге проверяют сразу три гипотезы о величинах задержки, отличающихся на $\tau_0/2$. Поэтому сдвиг фазы ПСП на каждом шаге должен составлять не более $1,5\tau_0$. Такой поиск продолжается до тех пор, пока хотя бы один тракт не зафиксирует появление свертки. Далее поиск прекращается, и система синхронизации по задержке переходит в режим слежения. В таком режиме работает дискриминатор величины ошибки, собранный на вспомогательных тракте и схеме

вычитания (см. рис. 10.5). Сигнал на выходе последней позволяет судить, на каком из вспомогательных трактов наблюдается свернутый по спектру радиосигнал большего уровня, т.е. на какой из трактов подается местная ПСП с фазой, более близкой к фазе ПСП сигнала. Эти данные позволяют выполнять подстройку фазы местной ПСП в правильном направлении, что делает работу системы синхронизации по задержке полностью аналогичной описанной выше работе систем тактовой синхронизации.

В заключение сделаем еще важное замечание по поводу работы описанной в настоящем пункте схемы синхронизации. Входящие в эту схему полосовые фильтры имеют полосы пропускания порядка битовой скорости передачи, осуществляемой с помощью ШПС. Хотя эта полоса во много раз (в базу раз) меньше полосы частот, занимаемых ШПС, она, как правило, недостаточна для эффективного подавления шумов, на фоне которых принимается радиосигнал. Если не принять дополнительные меры по их подавлению, то шумы будут существенно нарушать нормальную работу рассматриваемой системы синхронизации. В качестве такой меры не может быть использовано сокращение полосы пропускания ПФ, поскольку оно привело бы к подавлению части основных спектральных компонент свернутого (узкополосного) радиосигнала, так что отношение сигнал/шум не улучшилось бы. Поэтому остается только использовать последетекторную фильтрацию. В связи с этим во всех амплитудных детекторах описанной схемы должно использоваться последетекторное накопление; время такого накопления составляет длительность порядка 100 или, даже, нескольких сот элементарных посылок узкополосного сигнала.

10.2.6. Система синхронизации пакетов

Система синхронизации пакетов (ССП) представляет собой систему тактовой синхронизации, отличающуюся тем, что она не замыкается в каком-либо приемном устройстве, но распределена по всей системе связи. Ее можно описать на примере сотовой системы связи стандарта GSM. В данной системе каждый абонентский терминал (АТ) ведет передачу информации пакетами с длительностью каждого (длительность слота) 0,577 мс и периодом повторения пакетов (длительность кадра) 4,615 мс. Между пакетами одного терминала могут размещаться пакеты еще семи других таких терминалов, обеспечивая, за счет временного уплотнения, работу восьми каналов передачи на общей несущей частоте. Для этого необходимо так синхронизовать моменты излучения пакетов отдельными терминалами, чтобы на входе базовой станции (БС) их пакеты не перекрывались во времени. Именно выполнение этого требования и обеспечивается ССП. Ясно, что с помощью только устройств заключенных внутри АТ задача решена быть не может: терминал должен получать от БС сведения относительно правильности временного положения излучаемых терминалом пакетов и соответственно корректировать моменты их излучения. Можно считать, что терминал содержит в своем составе управляемый генератор местной последовательности моментов излучения пакетов (последовательность местных кадров, ПМК), которую он и передает на БС, где она подается на ФД, на другой вход которого подается эталонный сигнал, представляющий собой последовательность кадров БС (ПКБС). Вырабатываемые дискриминатором сигналы ошибки передаются на АТ, где проходя через соответствующее устройство усреднения (УУ), формируют сигнал управления ПМК.

Описанная структура полностью соответствует схеме кольца ФАП, приведенной на рис. 10.2. Здесь роль ОС играет ПКБС, а МС — ПМК. Особенность в данном случае заключается также в том, что цепи, соединяющие блоки (1 и 2) и (3 и 1) являются фактически радиолиниями достаточно большой протяженности и, следовательно, вносят весьма ощути-

мое запаздывание в транслируемые по ним сигналы. Так, при радиусе соты наземной сотовой связи порядка 30 км это запаздывание составит $\tau_p = 0,1$ мс, а в спутниковой связи достигает $\tau_p = 150$ мс. Следует иметь в виду, что в кольце фактически действует запаздывание, равное удвоенной величине времени распространения $\tau_3 = 2\tau_p$, поскольку сигнал проходит через две радиолинии (от 3 к 1 и от 1 к 2), фактически включенные последовательно. Соответственно, если через z обозначить запаздывание, измеренной в числе интервалов дискретизации $g = \tau_3/T_d$, то уравнение системы ФАП (10.1), преобразуется к виду

$$\theta_{j+1} = \theta_j - k_\phi \Phi(\psi_{j-g} - \theta_{j-g}) + 2\pi T_d f_j, \text{ где } f_{j+1} = f_j + k_f \Phi(\psi_{j-g} - \theta_{j-g}). \quad (10.13)$$

Результаты моделирования этой системы на ЭВМ представлены на рис. 10.6.

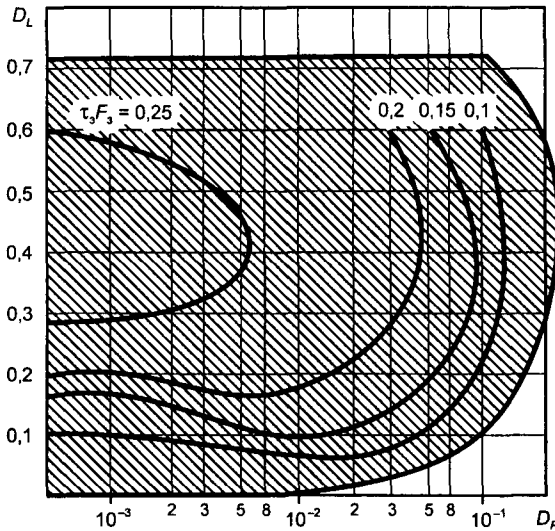


Рис. 10.6. Результаты моделирования работы ССП

В первую очередь цель моделирования — определение полосы захвата системы синхронизации при различных значениях параметров. Приведенные на рис. 10.6 результаты относятся к пилообразной характеристике фазового дискриминатора:

$$\Phi(x) = x/\pi \text{ при } |x| \leq \pi. \quad (10.14)$$

На рис. 10.6 представлена область устойчивости системы регулирования. Дело в том, что системы с запаздыванием всегда склонны к самовозбуждению и для того, чтобы это явление не возникало, необходимо делать инерционность системы достаточно большой, во много раз превосходящей τ_3 . При определении устойчивости использовались следующие параметры системы: $D_L = k_\phi \tau_3 / (2\pi T_d)$ и $D_f = k_f \tau_3^2 / T_d$. Область устойчивости лежит внутри кривой, охватывающей остальные кривые рис. 10.6. Сами кривые аналогичны линиям высоты на географической карте и соединяют точки, соответствующие одной и той же полосе захвата; параметром кривых рисунка является произведение $\tau_3 F_3$. Вообще говоря, кривые уровня представляют собой замкнутые, вложенные друг в друга кривые. Пересечение таких кривых, соответствующих различным значениям параметра $\tau_3 F_3$, невозможно, поскольку означало бы, что одной и той же комбинации параметров (D_L, D_f) соответствуют различные

значения полосы захвата. На рис. 10.6 приведена лишь части кривых уровня, поскольку моделирование производилось лишь для определенных сочетаний параметров.

Полученные данные позволяют сделать следующие выводы:

- введение регулирования частоты ($D_f > 0$) ни при каком выборе параметров не позволяет существенно увеличить полосу захвата системы; лишь при малых D_L удается увеличить ширину этой полосы, но не более чем в 1,5 раза. При этом D_f не должно превышать величин порядка 0,03–0,05; при больших значениях D_f полоса захвата быстро падает. Чем большее значение D_L используется, тем меньшим должно быть значение D_f ;
- наибольшая полоса захвата обеспечивается при $D_L \approx 0,5$ и $D_f = 0$. При этом остаточная ошибка по фазе отлична от нуля и может оказаться весьма значительной, что, как правило, не приемлемо. Поэтому целесообразно использовать управление частотой, что приведет к нулевым ошибкам по фазе, но сделать его столь малым, чтобы это не приводило к заметному снижению полосы захвата. Можно, например, рекомендовать параметры $D_L = 0,5$ и $D_f = 0,001$. При этом $F_3 \approx 0,25/\tau_3$.

Проведенное моделирование системы синхронизации с запаздыванием позволило оценить также длительности переходных процессов. Не приводя здесь полученные при этом точные данные, ограничимся лишь грубыми оценками. При $D_f = 0$ переходный процесс заканчивается обычно за время, равное $(3 - 5)\tau_3$, если параметры системы не слишком близки к границам области устойчивости (при $D_L < 0,65$). При введении регулировки по частоте, даже с небольшими рекомендованными выше значениями D_f , длительность переходного процесса резко возрастает до $(15 - 30)\tau_3$ и более.

Подводя итог, можно сказать, что система ФАП с запаздыванием, используемая обычно в качестве ССП, не позволяет обеспечить любое требуемое значение полосы захвата, какие бы параметры ФАП не выбирались. Максимально достижимое значение F_3 не превосходит $0,25/\tau_3$. Например, для систем спутниковой связи, использующих спутники-ретрансляторы (СР) на геостационарных орбитах (ГСО) с $\tau_3 \approx 0,3$ с, полоса захвата не может быть сделана с помощью рассматриваемых систем большей 0,9 Гц. Это обстоятельство, иногда, существенно ограничивает возможности построения соответствующих систем связи.

НЕКОТОРЫЕ ПЕРСПЕКТИВНЫЕ СИСТЕМЫ РАДИОСВЯЗИ

11.1. Спутниковые системы фиксированной службы радиосвязи и вещания

11.1.1. Искусственные спутники Земли в качестве ретрансляторов для систем связи

Задачи увеличения дальности и пропускной способности систем связи всегда были основополагающими проблемами данной области техники. К сожалению, соответствующие характеристики, как правило, оказываются альтернативными: мероприятия по увеличению пропускной способности приводят к сокращению дальности, и наоборот. В частности, повышение пропускной способности требует перехода на все более высокочастотные диапазоны волн, сигналы которых могут быть непосредственно переданы практически лишь на расстояния прямой видимости. Как средство разрешения этого противоречия, могут быть использованы ретрансляторы, поднятые достаточно высоко над поверхностью Земли.

Успехи развития космонавтики позволили использовать в качестве таких ретрансляторов ИСЗ. Поскольку они могут располагаться практически сколь угодно высоко над Землей, их область обслуживания может охватывать не только отдельные страны или моря, но и целые континенты и океаны. В общем случае спутники движутся по эллиптическим орбитам, в одном из фокусов которых располагается центр Земли. Спутник перемещается относительно наземного наблюдателя, а вместе с ним и область обслуживания перемещается по земной поверхности. В результате следует либо увеличивать число спутников в системе, либо согласиться с тем, что круглосуточная связь обеспечиваться не будет.

Улучшение ситуации может быть достигнуто, если орбиту спутника выбрать так, чтобы период обращения спутника вокруг Земли находился в простом соотношении с периодом ее обращения вокруг своей оси (синхронные орбиты). Использование таких орбит приводит к постоянному расписанию возможных сеансов связи, поскольку для любого наземного наблюдателя спутник-ретранслятор (СР) появляется в данной точке небесной сферы периодически, постоянно в одно и то же время.

Дальнейшие упрощения спутниковых систем связи наступают если:

- орбита спутника является круговой и лежит в плоскости экватора;
- период обращения спутника по орбите составляет ровно одни сутки.

Такой спутник вообще остается неподвижным относительно любого наземного наблюдателя. Соответствующая орбита именуется *геостационарной* (ГСО), а движущийся по ней спутник — *стационарным*. ГСО имеет радиус приблизительно 42,3 тыс. км. Она уникальная и единственная, поэтому размещение спутников на ней жестко контролируется между-

народными организациями во главе с действующим под эгидой ООН Международным союзом электросвязи (МСЭ). Той же организации поручена международная координация и других спутниковых систем связи с целью рационального ограничения взаимного влияния между ними.

Хотя в настоящее время подавляющая часть используемых СР являются стационарными, они не лишены существенных недостатков. Именно такие спутники лучше всего приспособлены для обслуживания тропических и субтропических регионов. По мере продвижения наблюдателя на поверхности Земли от подспутниковой точки вдоль меридиана к полюсам Земли, угол места направления на стационарный космический аппарат (КА) уменьшается, достигая нулевого значения для широты 82° (северной или южной). Для более близких к полюсам точек подспутникового меридиана видимость спутника вообще отсутствует. Легко понять, что граница геометрической видимости стационарного КА при отклонении наблюдателя от подспутникового меридиана опускается в направлении к экватору. Кроме того, работа радиолиний в направлениях с малыми углами места вообще резко затрудняется как за счет приема отраженных от Земли сигналов, так и за счет экранирующего действия различных возвышений, леса, строений или других препятствий. Поэтому стационарные КА практически неспособны обслуживать территории, лежащие севернее северного и южнее южного полярных кругов. Между тем эти территории часто представляют значительный интерес, например для России. Даже территория Северного полюса представляет значительный интерес, прежде всего в связи с тем, что через нее пролегают наиболее выгодные трассы ряда важнейших авиалиний.

Орбиты СР можно выбирать так, чтобы обеспечить преимущественное обслуживание тех или иных регионов на поверхности Земли. Так, в России была предложена эллиптическая орбита, специально приспособленная для обслуживания северных регионов нашей планеты. Апогей этой орбиты находится над северным полушарием на расстоянии приблизительно 40 тыс. км от поверхности Земли, а перигей лежит на высоте в несколько сотен километров над южным полушарием. Плоскость орбиты наклонена к экватору примерно на 65° . Период обращения спутника по этой орбите составляет половину суток, так что это синхронный спутник. За сутки он совершает два витка: первый из них, называемый *основным*, достигает апогея над Сибирью (в точке с географическими координатами $63,5^\circ$ с.ш. и 81° в.д.), а второй — сопряженный — в точке с той же широтой, но сдвинутой по долготе на 180° , т.е. 99° з.д. (над Канадой). Параметры этой орбиты выбраны так, что в примаыкающей к апогею части орбиты скорость углового перемещения спутника в направлении «восток–запад» совпадает с таковым для Земли. Это условие приблизительно выполняется на всем рабочем участке орбиты (от трех-четырех часов до достижения апогея до трех-четырех часов после его прохождения) и обеспечивает отсутствие перемещения спутника по отношению к любому наблюдателю на Земле в направлении «восток–запад».

На рабочем участке орбиты сравнительно небольшим оказывается и перемещение в направлении «север–юг». Эллиптическая орбита обеспечивает обслуживание северного полушария Земли, включая и область Северного полюса с достаточно большими углами места. Недостатком ее является необходимость использования системы из трех-четырех спутников для поддержания непрерывности связи в течение суток, что удорожает космический сегмент системы; также существенно, что при использовании эллиптических спутников на ЗС приходится обеспечивать слежение антенной за перемещениями КА, что удорожает и земной комплекс системы.

Спутник-ретранслятор (СР) должен принимать сигналы от земных станций (ЗС) системы связи, усиливать их и вновь передавать на те ЗС, которым они предназначены. Таким образом, СР содержит приемное и передающее оборудование для ретрансляции сигналов.

Поскольку сквозное усиление приемопередающего тракта СР должно быть достаточно большим, необходимо вести прием и передачу на разных частотах (в противном случае не удастся избежать самовозбуждения тракта). Таким образом, обязательным элементом тракта ретрансляции являются также преобразователи частоты.

Особенность ретрансляторов вещательной службы в том, что для них основным является передающий тракт, через который собственно и осуществляется вещание. На вещательных СР устанавливается и приемное оборудование, используемое для приема подаваемых на борт вещательных программ. Радиолиния подачи программ на борт именуется *фидерной*.

Спутник-ретранслятор, как всякий активный КА, кроме собственно тракта ретрансляции, именуемого по отношению к этому аппарату, *полезной нагрузкой* (ПН), содержит также и целый ряд вспомогательных систем, таких как система электропитания, система ориентации и стабилизации, система терморегулирования и управления. Последняя включает системы формирования и передачи телеметрической информации. КА за вычетом полезной нагрузки именуется *космической платформой* (КП). Такая платформа может использоваться в сочетании с различными ПН для создания ряда различных КА.

В настоящее время в интересах фиксированной и вещательной служб чаще всего используются стационарные СР. Типовые параметры платформ таких спутников:

- энерговооруженность до 5–7 кВт, причем для питания полезной нагрузки выделяется 1,5–2 кВт;
- масса порядка 2–3 т, в том числе полезной нагрузки 0,5–0,8 т;
- точность ориентации и стабилизации порядка $0,1^\circ$;
- срок активного существования 12–15 лет.

Наряду с типовыми КА в настоящее время считается перспективным использование в интересах фиксированной службы малых КА (МКА) с массой 500–800 кг (в том числе ПН 100–200 кг) и энерговооруженностью 1,8–2,5 кВт. Достоинство МКА — возможность группового или попутного (вместе с типовым КА) запуска, что существенно снижает расходы на выведение. МКА могут запускаться в те точки, где уже расположены другие СР и обеспечивать необходимое дополнение работающих или на них стволов или замену стволов, вышедших из строя. На них могут строиться также национальные системы спутниковой связи сравнительно небольших или небогатых стран.

В зависимости от состава пользователей СР делятся на международные и национальные. Наиболее известные международные СР фиксированной службы Intelsat и Eutelsat. Существенными ресурсами владеет также международная компания Интерспутник. СР Eutelsat содержат также стволы, чаще всего используемые европейскими странами для телевизионного вещания. Специально для этих целей используется спутниковая система Astra.

Национальная система спутниковой фиксированной службы России в настоящее время использует СР типа «Экспресс», а также «Ямал» различных модификаций.

11.1.2. Диапазоны частот и зоны обслуживания

Для систем фиксированной спутниковой службы Регламентом радиосвязи выделен ряд диапазонов частот; эти диапазоны принято обозначать буквами или в виде дроби, числитель которой грубо оценивает рабочие частоты для линии от СР на ЗС, а знаменатель — для линии от ЗС на СР. В интересах фиксированной службы выделены диапазоны: С — 4/6 ГГц, Ku — 11/14 ГГц и Ka — 20/30 ГГц или 20/40 ГГц. Выделен также участок X диапазона 7/8 ГГц, который, по негласному соглашению между странами используется только для военной связи. В каждом из указанных диапазонов Регламентом радиосвязи в интересах фиксиро-

ванной спутниковой службы выделена полоса рабочих частот 500–1000 МГц как для линии вверх (от ЗС на СР), так и для линии вниз (от СР на ЗС).

Что касается радиовещания, то для спутникового аудиовещания выделена полоса в районе 1,4 ГГц, а для телевизионного вещания в качестве типового принимается диапазон 12 ГГц (полоса порядка 800 МГц) при работе фидерной линии в диапазоне 18 ГГц.

Для спутниковой связи характерны зоны обслуживания (ЗО), охватывающие целые регионы, или континенты, или, даже, почти всю поверхность Земного шара. Соответственно по ЗО системы делят на региональные и глобальные. Глобальное обслуживание реализуют чаще всего международные спутниковые системы связи. Для национальных систем более характерно региональное обслуживание. Часто на СР реализуются как антенны, освещающие всю видимую с КА территорию Земли, так и освещающие лишь отдельные регионы. В результате часть связного ресурса данного КА может использоваться в интересах глобальной связи, а часть — региональной.

11.1.3. Структура и характеристики полезной нагрузки

Поскольку каждый СР используется, как правило, целым рядом ЗС, проблема многостанционного доступа в спутниковой связи считается одной из основных. В системах фиксированной и вещательной служб чаще всего реализуется частотное деление (см. 6.1). Практически во всех случаях оборудование ПН выполняется в виде ряда параллельно действующих стволов ретрансляции с неперекрывающимися полосами пропускания. Типовые полосы стволов составляют 30–40 МГц или 72 МГц.

Каждый ствол имеет свой передатчик (усилитель мощности — УМ). Часто несколько стволов используют общую антенну. При этом они подключаются к ней через систему соответствующих СВЧ фильтров (мультиплексер), исключающих влияние УМ различных стволов друг на друга, а также исключающих излучение стволом помех (гармоник несущей частоты, комбинационных составляющих) на частотах, лежащих вне полосы пропускания данного ствола.

Что касается приемных трактов, то каждый из них применяется в интересах нескольких стволов, так что их полоса пропускания измеряется величинами в несколько сотен МГц. Каждый приемник содержит преобразователи частоты, на выходах которых специальными фильтрами выделяются полосы частот для каждого из стволов СР; к выходам этих фильтров и подключаются передающие тракты стволов. В типовых вариантах в настоящее время используется однократное преобразование частоты, сдвигающее частоту приема на СР (частота линии вверх) на частоту передачи с СР (линии вниз).

Типовые СР в настоящее время включают 20–30 стволов, как правило, часть в С-, а часть в Ки-диапазонах. Проектируются и МКА с числом стволов до 10. УМ типовых СР строятся на лампах бегущей волны (ЛБВ). Достоинством этих приборов, особенно существенным при их применении на КА, является высокий КПД в режиме насыщения (до 50–60%). Их недостаток связан с необходимостью значительного снижения выходной мощности для перехода в линейный режим (порядка 6 дБ). Правда, в последнее время для УМ на ЛБВ разработаны специальные амплитудные корректоры, позволяющие уменьшить требуемое для линеаризации усилителя снижение мощности до 2–3 дБ. Эти устройства сложны в настройке и поэтому достаточно дороги. Кроме того, ЛБВ имеют значительные массу и габарит и требуют высоковольтного питания, что затрудняет конструирование. В связи с этим постоянно ведутся работы по созданию твердотельных УМ в диапазонах фиксированной службы. Основная задача — повышение КПД до соизмеримых с ЛБВ значений. На данном пути имеются значительные успехи. Уже в настоящее время твердотельные

усилители иногда применяют в С-диапазоне в тех случаях, когда основным признается линейный режим работы ствола. В этих условиях твердотельные усилители обеспечивают практически тот же КПД, что и УМ на ЛБВ при меньшей массе и габаритах, причем они не требуют высоковольтного питания.

Мощность УМ в режиме насыщения составляет, как правило, для СР фиксированной службы величину порядка 15 Вт, что при глобальной ЗО соответствует эквивалентной изотропно-излучаемой мощности (ЭИИМ) ствола в направлении на любую точку ЗО (в углах связи) не ниже величины порядка 25 дБВт. При региональной ЗО типовое значение ЭИИМ возрастает до 32–38 дБВт в зависимости от размеров зоны. Для СР телевизионного вещания в направлении на абонентские терминалы (АТ) используются УМ с мощностью в насыщении до 100–120 Вт. При этом всегда используется региональное обслуживание и ЭИИМ ствола составляет 50–55 дБВт.

11.1.4. Многостанционный доступ

Из изложенной типовой структуры ПН следует, что здесь используется многостанционный доступ с частотным делением между стволами. Часто такое деление дополняется еще делением по направлению прихода луча и по поляризации (см. 6.4). Например, если СР должен обслуживать несколько в достаточной степени географически разнесенных регионов, то можно, обслуживая каждый такой регион отдельным лучом бортовых антенн, и повторять рабочие частоты в различных регионах. Это означает, что мы заменяем частотное разделение, делением по углу прихода волны.

Для обеспечения повторного использования полос рабочих частот часто применяется поляризационное разделение. При этом обеспечивается параллельное функционирование совпадающих по частоте стволов на двух ортогональных поляризациях. С этой целью в составе ПН создаются приемные и передающие антенны на каждую из ортогональных поляризаций. Пара таких антенн использует общее зеркало, но отдельные облучатели (каждый на свою поляризацию). Все остальные элементы трактов стволов для каждой поляризации свои.

За счет пространственных методов разделения в СР Intelsat-8 удастся достичь почти четырехкратного повторения рабочих частот. Следует отметить, что в случае поляризационной развязки приходится предъявлять весьма жесткие требования к коэффициентам эллиптичности K_z приемных и передающих антенн СР и ЗС. Так, в соответствии со стандартом Intelsat от всех антенн требуется $K_z > 0,93-0,96$.

Выше охарактеризованы типовые способы разделения сигналов между стволами. Остановимся теперь на типовых способах многостанционного доступа к ресурсу каждого ствола. Чаще всего здесь имеет место частотное разделение, поскольку оно не требует какой-либо синхронизации работы ЗС, использующих общий ствол ретрансляции. Достаточно, чтобы полосы частот, занимаемых сигналами этих ЗС, не перекрывались и лежали внутри полосы пропускания ствола. Это позволяет операторам продавать связные ресурсы ствола по частям, имея ввиду их использование теми ЗС, которые есть у потребителя, не заботясь об их унификации.

Отметим, что часто в системах фиксированной службы сообщения, транслируемые каждой ЗС, являются многоканальными. Каждая ЗС, будучи установленной у некоторого потребителя, обеспечивает его одновременно различными видами связи, как, например, телефонной связью (возможно по нескольким каналам), передачей данных, работой в сети Интернет и т.п. Каждый канал может быть адресован своему корреспонденту и, следовательно, возможно различным ЗС сети. В общем случае, для реализации такой схемы соединений каж-

дая ЗС должна принимать несколько радиосигналов на различных несущих от тех станций сети, с которыми она поддерживает связь. Это усложняет оборудование ЗС: в приемнике теперь должно формироваться несколько трактов приема, каждый со своим преобразователем, трактом ПЧ, демодулятором, декодером и др.

При многоканальной передаче на каждой ЗС должно обеспечиваться уплотнение транслируемых сообщений. С этой целью используют один из двух следующих вариантов:

- временное уплотнение (см. 6.1);
- частотное уплотнение с передачей сообщения каждого канала на своей несущей частоте. Этот вариант обозначают как один канал на несущую (ОКН).

Вариант с временным уплотнением гораздо проще реализуется на передающей стороне. Однако требует установки на приемной стороне в каждом тракте приема отдельной аппаратуры временного разделения, часто и асинхронного. Наоборот, ОКН гораздо сложнее реализуется на передаче, но исключает использование какого-либо оборудования разделения каналов на приеме. Успехи цифровой техники приводят к тому, что в настоящее время вариант с временным уплотнением оказывается в целом предпочтительнее. Вариант ОКН находит применение лишь в отдельных специальных случаях.

Недостатком частотного разделения внутри ствола, к которому, в частности, приводит и вариант с ОКН, является возникновение комбинационных искажений в УМ и необходимость недоиспользования УМ по мощности (см. 6.2). Поэтому в тех случаях, когда синхронизация работы всех ЗС, использующих данный ствол, оказывается возможной, большая эффективность может быть достигнута при выборе внутри ствола только временного уплотнения. В этом случае каждая ЗС, работающая через рассматриваемый ствол, группирует передаваемую информацию в периодически (по времени) следующие пакеты. В интервале между пакетами данной ЗС передаются пакеты других ЗС. Чтобы пакеты отдельных ЗС не перекрывали друг друга на входе ретранслятора, работа ЗС должна быть синхронизована во времени с учетом различия расстояний от этих станций до СР. Принципы построения соответствующих ССП описаны в гл. 10. Частота следования пакетов от каждой ЗС определяется ведущей станцией рассматриваемой сети и, в общем случае не синхронна с тактовой частотой цифровых сообщений, подлежащих передаче от каждой из ЗС данной сети. Это означает, что на ЗС должен осуществляться асинхронный ввод цифровых сообщений в передаваемый цифровой сигнал, как это было описано в 6.3.

При использовании в стволе временного уплотнения (пакетной передачи) в любой момент времени через ствол транслируется не более, чем один сигнал (один пакет). Если используемые сигналы имеют неизменную амплитуду, то УМ ствола может работать в режиме насыщения, т.е. с максимальной ВЧ-мощностью, так что энергетический выигрыш от перехода на временное уплотнение (от частотного) оказывается максимальным. Однако для повышения спектральной эффективности иногда используют методы, приводящие к появлению амплитудной модуляции сигнала (см. 6.2). При этом выигрыш от временного уплотнения сокращается.

11.1.5. Структура и характеристики земных средств

В типовых случаях ЗС имеют зеркальные прямо-передающие антенны. В приемных трактах, как правило, реализуются твердотельные маломощные усилители (МШУ), устанавливаемые прямо на антенне. Там же устанавливается и прямо-передающий конвертор (преобразователь) с перестраиваемыми гетеродинами, позволяющий выбирать стволы ретрансляции по приему и передаче.

Антенна с установленным на ней оборудованием выносится либо на специальную площадку, либо на крышу здания. Остальное оборудование станции размещается в помещениях и носит название *кабинетного*. Оно содержит сравнительно низкочастотные тракты промежуточной частоты (ПЧ), демодуляторы и декодеры, модуляторы и кодеры, а при необходимости аппаратуру уплотнения и разделения. Что касается УМ, то его тоже выносят на антенну. Соединение оборудования антенного поста с кабинетным осуществляется по достаточно протяженному (обычно до 200 м) кабелю.

В сетях фиксированной службы используют ЗС разного уровня. Последний определяется числом поддерживаемых ею каналов и направлений связи. Соответственно изменяются энергетические параметры станции (диаметр зеркала, мощность передатчика), и объем кабинетного оборудования. Выделяют центральные станции (ЦЗС), поддерживающие связи с рядом основных станций системы. В каждой системы таких станций бывает одна или несколько. Как правило, на ЦЗС располагается комплекс оборудования системы автоматизированного управления связью (АСУ-С), использующего каналы управления и контроля, организуемые ЦЗС. В связи с ведущей ролью ЦЗС в системе в ней предусматривается в более широком масштабе резервирование аппаратуры, включая даже антенны. Иногда ЦЗС работает сразу по нескольким СР, так что в ней может одновременно использоваться несколько антенн.

Далее, ЗС, поддерживающие связь с несколькими другими ЗС сети, т.е. по нескольким направлениям, часто именуют *узловыми* в отличие от оконечных, работающих только по одному направлению.

На первоначальном этапе создания спутниковой фиксированной службы использовались ЗС с достаточно большими антеннами (диаметр зеркала 12 и, даже, 30 м). В настоящее время типовые ЗС таких систем имеют антенны с диаметром зеркала порядка 5–9 м. Более того, развивается системы с небольшими ЗС (класса Very Small Aperture Terminal, VSAT), с диаметром антенн 1,8–3,5 м. ЗС этого класса размещаются непосредственно у потребителей, имеют передатчики небольшой мощности (1–20 Вт), не требуют постоянного технического обслуживания и контролируются и управляются централизованно.

Что касается абонентских терминалов телевидения (АТТ), то в настоящее время они, как правило, являются только приемными. Диаметр зеркала антенны этих ЗС находится в интервале 0,6–1 м. Такие терминалы устанавливаются у потребителя и, кроме антенного, включают еще кабинетный блок, выполняющий функции выбора ствола ретрансляции, демодуляции, декодирования, временного разделения и выбора программы для просмотра. Сам просмотр осуществляется на домашнем телевизоре абонента. Часть передаваемых по спутниковым каналам программ в типовом случае подвергается специальному преобразованию, так что их несанкционированный просмотр оказывается невозможным. Для реализации возможности их приема абонент должен внести специальную плату, после чего его терминалу передается специальный ключ, обеспечивающий возможность вскрытия программы.

11.1.6. Тенденции развития и совершенствования

Основной проблемой систем спутниковой связи вообще и фиксированной службы и радиовещания, в частности, является перегрузка используемых диапазонов частот. Так, в С- и Ku-диапазонах число поданных с МСЭ запросов на предоставление рабочих частот для КА на геостационарной орбите (ГСО) во много раз превосходит имеющиеся ресурсы. Правда, Ка-диапазон в настоящее время используется далеко не полностью и представляет основной резерв дальнейшего развития данного направления техники связи. Следует иметь в виду, что использование данного диапазона сопряжено со значительными техническими трудностями.

Прежде всего, переход на этот диапазон приводит к значительным энергетическим потерям, связанным, прежде всего, со значительным поглощением соответствующих радиоволн во влаге атмосферы и дождях. Величина этих потерь существенно зависит от погодных условий в месте расположения ЗС и протяженности участка спутниковой трассы, проходящего через атмосферу.

Для компенсации этих потерь приходится реализовывать в спутниковых радиолиниях дополнительные энергетические запасы, величина которых зависит еще от требуемой абонентами надежности функционирования связи. В общем случае, по этим причинам трудно указать сколько-нибудь точное значение необходимых дополнительных запасов. Ориентировочно они составляют 6 дБ на частотах порядка 20 ГГц и 13 дБ на 40 ГГц. Следует также иметь в виду, что в Ка-диапазоне значительно возрастают потери в кабелях, волноводах, соединителях, ферритовых изделиях, фильтрах и т.п. В результате энергетические потери при переходе в Ка-диапазон оказываются столь значительными, что возникает вопрос об их компенсации путем радикального изменения принципов обслуживания путем перехода к сотовым структурам (см. 6.4). Однако поскольку в настоящее время такие решения используются в основном подвижной спутниковой службой, соответствующие возможности будут рассмотрены в 11.3.

Что касается спутниковых систем телевизионного вещания, то основное достижение последних лет в этой области — переход на цифровую передачу изображений со сжатием спектра. Ранее телевизионный сигнал передавался по спутниковым радиолиниям в аналоговом виде с использованием частотной модуляции. При этом ретрансляция одной программы требовала как по занимаемой полосе частот, так и по мощности выделения ствола с полосой не менее 30 МГц и мощностью порядка 100 Вт. Даже через самые мощные из действующих СР телевещания могли обеспечить трансляцию лишь десятка программ, что соизмеримо с числом программ в крупных городах, создаваемых наземными средствами. Это обстоятельство значительно снижало интерес пользователей к спутниковому вещанию.

Между тем специалистам давно было ясно, что телевизионный сигнал содержит весьма значительную избыточность, хотя бы потому, что передача неподвижных элементов изображения, составляющих часто большую часть информации, многократно (с частотой кадров) повторяется. Избыточной телевизионная развертка является и для передачи хотя и движущихся элементов изображения, но с низкой динамикой. Поэтому с середины XX в. велись работы по сжатию телевизионного сигнала. Основная трудность заключалась здесь в высокой чувствительности глаза к малейшим нарушениям естественности изображения. И лишь в 1990-х гг. ее удалось преодолеть. Разработанный стандарт позволил сократить объем цифрового представления телевизионного сигнала до 2–8 Мбит/с. Нижняя граница соответствует передаче с низкой динамикой (передачи из студии новостей, интервью и т.п.), а верхний — с высокой динамикой (спортивные соревнования). Эффективность такого стандарта становится очевидной, если сопоставить приведенные объемы сжатых ТВ сообщений с их объемами, получающимися без сжатия, например, путем ИКМ-преобразования. Последние составляют порядка 100 Мбит/с.

Использование этого стандарта позволяет за счет того же ретрансляционного ресурса, который раньше выделялся для трансляции одной телепрограммы в аналоговом виде, осуществлять передачу 6–8 программ в цифровой форме. Программы уплотняются по времени, образуя групповой сигнал со скоростью примерно 35 Мбит/с и передаются через один ствол на общей несущей. В кабинетном блоке АТТ по желанию пользователя выделяется ствол и после демодуляции и временного разделения — требуемая программа. В результате через один СР можно организовать 60–80 телевизионных каналов, что резко повысило интерес пользователей к спутниковому вещанию.

Выше были описаны АТТ, не имеющие передающего оборудования. В последнее время с развитием услуг мультимедиа, начал возрастать спрос на интерактивное телевидение, когда абоненту предоставляется возможность запроса необходимой телевизионной программы. Эта услуга хорошо сопрягается с задачей создания высокоскоростных каналов доступа абонентов к сети Интернет. На первом этапе перечисленные услуги обеспечиваются следующим способом: абонент посылает свой запрос на услугу (телевизионную программу, сведения из сети Интернет) по проводным каналам телефонной сети общего пользования, через обычный телефонный модем, сопряженный с его персональным компьютером. Ответное сообщение он получает через свой АТТ.

С этой целью одна или несколько программ телевидения заменяются соответствующей информацией. В таком варианте полезно используется то обстоятельство, что мультимедийные линии связи, как правило, несимметричны: объем информации в направлении от абонента на центр (запросы) оказывается значительно меньше объема информации, передаваемой в обратном направлении (ответы центра).

Однако такое решение может быть приемлемым лишь на начальном этапе. В настоящее время в ряде систем предусматривается использование АТТ с передающим трактом, что исключает необходимость применения проводной связи при организации мультимедийной сети. Следует отметить, что такие ЗС следовало бы, строго говоря, относить уже к фиксированной спутниковой службе; во всяком случае, здесь объединяются задачи данной службы и телевидения. Создание соответствующих АТТ сопряжено с двумя основными проблемами:

- введение передающего тракта не должно приводить к резкому удорожанию АТТ;
- наличие такого тракта не должно приводить к необходимости получения специального разрешения на установку каждой такой станции.

Последнее требование нуждается в дополнительном пояснении. Дело в том, что установка приемо-передающих станций спутниковой связи в каждом конкретном случае требует специального разрешения ряда государственных и муниципальных органов власти. Здесь учитываются и проблемы экологии, и возможности здания, где устанавливается ЗС и т.п. Ясно, что такая процедура, если ее распространить на АТТ, сделает практически невозможной развитие соответствующей сети. (При только приемных АТТ небольшого размера данная проблема, естественно, не возникает.) Аналогичные трудности при продаже сотовых телефонов имели бы место и в системах сотовой связи (наземных и спутниковых). Однако с самого начала было установлено, что в диапазонах волн, выделенных для подвижной связи, нет необходимости получать разрешение на каждый отдельный абонентский терминал (АТ); достаточно сертифицировать в установленном порядке АТ данного типа. Таким образом получается, что для интерактивных АТТ следует либо использовать на передачу диапазоны рабочих частот выделенные для подвижной связи, либо добиваться выделения специального участка диапазона для запросных каналов. Первый путь может вызывать возражения, поскольку АТТ это станции скорее фиксированной службы, чем подвижной. Представляется, что второй путь более правилен, по крайней мере, юридически.

11.1.7. Энергетический расчет спутниковых радиолиний

Энергетические расчеты спутниковых радиолиний осуществляется с использованием соотношений, приведенных в 1.1. Необходимо учитывать, что речь идет о радиолинии с ретрансляцией. Кроме того, должны быть конкретизированы величины потерь, применительно к радиолиниям данного типа. Для определенности, ограничимся выражениями, относящимися к случаю реализации стационарных СР.

Особенности расчетов радиолиний с ретрансляцией связаны с необходимостью учета шумов приемной системы ретранслятора (для краткости — шумов ретранслятора). В спутниковых радиолиниях такие шумы влияют на общую энергетику двояким образом: *во-первых*, усиливаясь стволом вместе с полезными сигналами, они отбирают на себя часть мощности УМ, снижая соответственно ЭИИМ полезных сигналов; *во-вторых*, будучи ретранслированы, шумы ретранслятора воздействуют на вход приемной системы ЗС, увеличивая ее уровень шума.

Дополнительные, связанные с ретрансляцией особенности, возникают в тех типовых случаях, когда через ствол транслируется множество сигналов с частотным делением. Поскольку выходная мощность УМ ствола ограничена и так или иначе перераспределяется между всеми ретранслируемыми им сигналами, то парциальный ЭИИМ ретранслятора, приходящийся на данную несущую, зависит от уровней остальных несущих. Ниже мы рассмотрим два типовых случая:

- ствол ретранслирует множество сигналов, но работает в квазилинейном режиме (КЗЛР);
- ствол усиливает в каждый данный момент времени один сигнал и работает в режиме насыщения (РН).

В первом случае ствол характеризуется коэффициентом усиления K , измеряемым от выхода приемной антенны до входа передающей. Отметим, что в современных ретрансляторах эта величина допускает регулировку в широких пределах (15–20) дБ по командам с Земли. Для расчета во втором случае используют, задаваемый в спецификации на ПН, уровень сигнала на выходе приемной антенны $p_{сн}$, обеспечивающий перевод УМ ствола в режим насыщения.

Расчеты спутниковых радиолиний базируются на общем соотношении (1.16). Поскольку эти линии практически всегда работают в режиме прямой видимости, следует принять $\beta = \nu = 1$. Далее, для СР на ГСО максимальное расстояние от любой точки ЗО на поверхности Земли не превосходит величины $R = 40$ тыс. км. Подставляя эти величины, получим для соотношения сигнал/шум на выходе приемной антенны спутниковой радиолинии:

$$H \text{ [дБ]} = 14 - \eta \text{ [дБ]} + P \text{ [дБВт]} + U \text{ [дБ/К]} - 20 \lg f \text{ [ГГц]} - 10 \lg \Delta f \text{ [кГц]}. \quad (11.1)$$

Соотношение (11.1) используют как базовое при расчете спутниковых радиолиний. Ниже описан порядок такого расчета применительно к некоторой несущей, транслируемой ретранслятором.

Как правило, расчет начинают с линии вверх. Используя (11.1) и подставляя вместо P ЭИИМ данной несущей на выходе передающей ЗС P_3 , а вместо U приведенную добротность приемной системы СР, найдем отношение этой несущей на входе ретранслятора к шумам его приемной системы H_p . В качестве ЭИИМ и приведенной добротности подставляют их значения на направлении от СР на передающую ЗС, рабочая частота и потери должны соответствовать линии вверх, а полоса пропускания должна определяться в соответствии с шириной манипуляционного спектра транслируемой несущей. Далее с использованием (1.15а) вычисляют мощность принятого сигнала на выходе приемной антенны СР:

$$p_c \text{ [дБВт]} = p_w \text{ [дБВт]} + H_p \text{ [дБ]}, \quad (11.2)$$

где p_c — мощность шума ретранслятора в полосе Δf , приведенного к выходу приемной антенны.

Следующий этап расчета относится к прохождению принятой несущей через ствол ретрансляции. При работе в режиме КЗЛ парциальная ВЧ-мощность на входе передающей антенны ствола, приходящаяся на данную несущую, выражается соотношением

$$P_c \text{ [дБВт]} = p_c \text{ [дБВт]} + K \text{ [дБ]}. \quad (11.3)$$

Прибавляя к этой величине коэффициент усиления передающей антенны СР (дБ) в направлении на принимающую ЗС, найдем парциальный ЭИИМ ретранслятора P_p для рассматриваемой несущей в данном направлении.

Если речь идет о работе в режиме насыщения, то параметр P_p берется из спецификации на ПН. Однако для того, чтобы он был фактически реализован, необходимо обеспечить предусмотренный этой спецификацией уровень сигнала на входе ствола. Это требование в соответствии с (11.2) определяет минимально допустимое значение H_p , которое, в свою очередь, позволяет вычислить с использованием (11.1) требуемое значение ЭИИМ ЗС P_s .

Теперь можно переходить к расчету радиолинии вниз. Вначале находят отношение сигнал/шум H_s на выходе приемной антенны принимающей ЗС без учета шумов борта. Эту величину получают при подстановке в (11.1) значений P_p и приведенной добротности принимающей ЗС U_s и значений остальных величин, соответствующих линии вниз. Далее необходимо учесть еще влияние шумов СР, сопровождающих ретранслированный сигнал. С учетом шумов на выходе приемной антенны принимающей ЗС будет действовать шум, представляющий собой сумму двух компонент: шумов приемной системы этой ЗС и ретранслированных шумов СР. Мощность первой из них меньше мощности принимаемого полезного сигнала в H_s раз, а мощность второй — в H_p раз. Поэтому отношение мощности полезного сигнала на входе принимающей ЗС к мощности суммарных шумов составит

$$H_0 = \frac{H_p H_s}{H_p + H_s} = H_s \frac{H_p/H_s}{1 + H_p/H_s} = H_s/\chi, \quad (11.4)$$

где величина $\chi = 1 + H_s/H_p$ учитывает потери, вызванные воздействием ретранслированных шумов СР. Следует подчеркнуть, что в последние соотношения отношения мощностей сигнала и шума H должны подставляться в разгах, а не в дБ.

Отметим, что фактически соотношение (11.4) учитывает накопление шумов линий вверх и вниз в общем тракте с ретрансляцией. Зависимость потерь на такое накопление от превышения H_p над H_s , определяемой величиной Δ [дБ] = $10 \lg \{H_p/H_s\} = H_p$ [дБ] - H_s [дБ] представлена в табл. 11.1.

Таблица 11.1. Потери, вызванные накоплением шумов

Δ , дБ	-10	-5	-3	0	2	4	6	8	10	12
χ , дБ	10,4	6,2	4,8	3	2,1	1,5	1	0,64	0,4	0,27

Большие значения потерь указывают на то, что линия вверх является энергетически гораздо более напряженной, чем линия вниз, так что именно линия вверх и определяет помехоустойчивость передачи. Наоборот, малые потери означают, что именно линия вниз фактически характеризует работу сквозного тракта передачи сигнала. В спутниковой связи гораздо труднее оказывается обеспечить требуемую энергетику именно в линии вниз, в связи с чем, как правило, осуществляется работа с малыми потерями χ .

В заключение параграфа остановимся еще на реальных значениях потерь η , входящих в выражение (11.1), которые должны учитываться в расчетах спутниковых систем фиксированной службы. Источником таких потерь служат следующие причины: замирания сигнала, вызванные многолучевым распространением, поглощение радиоволн на трассе распространения, потери из-за неточного наведения антенн радиолинии, потери, связанные с эллиптичностью поляризации используемых в радиолинии электромагнитных волн и т.п. Первый из перечисленных источников потерь в линиях спутниковой фиксированной службы не действует, поскольку связь, как правило, осуществляется прямым лучом, а высокая направленность антенн исключает многолучевой прием. На первое место по значимости для таких радиолиний выступает поглощение радиоволн во влаге атмосферы и дожде. Если для С-диапазона эти потери можно не учитывать, то уже для Ку-диапазона они оказываются основными. Для самой грубой оценки их величины можно считать, что, будучи выражены в дБ, они линейно нарастают от частоты 6 ГГц, где они еще отсутствуют ($\eta = 0$ дБ), до значения $\eta = 13$ дБ (см. выше) на частоте 40 ГГц. Все остальные источники приводят суммарно к потерям, не превосходящим 1–1,5 дБ.

11.2. Сотовые наземные системы связи

11.2.1. Общие сведения

Последние десятилетия ознаменовались резким прогрессом в развитии систем подвижной службы, прежде всего, за счет реализации сотового принципа обслуживания подвижных абонентов (см. 6.1.4). Напомним, что суть этого принципа заключается в том, что вся требуемая область обслуживания (ОО) разбивается на парциальные зоны, именуемые *сотами*, и связь с подвижными абонентскими терминалами (АТ) каждой соты поддерживает своя приемно-передающая радиостанция. Последние называются *базовыми станциями* (БС).

Базовые станции всех сот соединены линиями связи с центром коммутации (ЦК) системы, который управляет их работой, обеспечивает при необходимости ретрансляцию информации между ними, а также привязку к основным наземным сетям общего пользования, таким как телефонная сеть, сеть передачи данных и т.п. В результате система может обеспечивать связь как между АТ, так и АТ с абонентами наземных сетей общего пользования.

Во всех случаях передача информации в сетях сотовой связи обеспечивается от (на) АТ путем ретрансляции через БС той соты, в которой данный АТ размещается во время сеанса связи. Если необходимо установить соединение между двумя АТ, расположенными в разных сотах, то осуществляется ретрансляция через обе БС этих сот.

Широкое распространение получили цифровые наземные системы сотовой связи двух типов: Global System for Mobile Communications (GSM) и CDMA. Принципиальное отличие данных систем заключается в используемых методах многостанционного доступа: если CDMA, как это следует из самого названия, использует широкополосные сигналы и их разделение по форме, то GSM реализует узкополосные сигналы и их частотно-временное разделение. Ниже каждая из этих систем будет кратко описана.

Предварительно остановимся на некоторых общих для всех систем сотовой связи процедурах. Прежде всего, отметим, что каналы связи здесь всегда предоставляются только по требованию абонентов. При возникновении потребности в связи, пользователь, передает соответствующий запрос в ЦК с указанием как своего идентификационного номера, так и номера вызываемого абонента. При этом в качестве пользователя, инициирующего организацию канала связи, может выступать как АТ, так и любой абонент наземной сети общего пользования, с которой сопряжена данная система сотовой связи. Передача запроса с АТ, расположенного в некоторой соте, осуществляется через БС этой соты по специальному запросному радиоканалу (ЗРК); такие каналы создаются в каждой соте.

Сведения о радиоданных этих каналов передаются через специальные «широковещательные» каналы (ШРК), также создаваемые в каждой соте. И ШРК, и ЗРК это симплексные каналы; первые транслируют информацию только от ЦК к АТ (данное направление принято именовать прямым), а вторые, наоборот, от АТ к ЦК (обратное направление). ШРК используется также для передачи на АТ команд управления, представляющих собой реакцию ЦК на поступивший запрос. Соответствующие команды посылаются как в соту, где находится инициирующий связь АТ, так и в виде вызова в соту размещения вызываемого АТ или в наземную сеть, если вызывается абонент этой сети.

В результате либо устанавливается требуемое дуплексное соединение, либо запрашивающий абонент получает отказ по причинам отсутствия свободных каналов, занятости вызываемого абонента, его недоступности, отсутствия денежных средств на счете вызывающего и т.п.

В действительности, кроме ШРК и ЗРК в системах сотовой связи используется и ряд других радиоканалов управления, уже, как правило, дуплексных. Использование этих каналов либо предшествует сеансу связи, либо сопровождает его. В последнем случае (сопрово-

ждающий канал управления) служебная информация объединяется с основной методом временного уплотнения. Полное описание используемых каналов управления выходит за рамки задач настоящего раздела. Соответственно, алгоритмы функционирования радиолиний сотовой связи здесь описываются лишь в принципиальном плане; детали соответствующих процедур опущены.

Важную роль в алгоритмах функционирования сотовых сетей играют данные о соте, в которой размещаются те или иные АТ-системы. Для хранения таких данных на ЦК используется два регистра памяти: так называемый *домашний регистр*, содержащий данные об АТ, приписанных к ЗО данного ЦК, и гостевой регистр с данными о появившихся в этой зоне АТ, приписанных к другим ЦК, с которыми предусмотрен роуминг. В тех же регистрах хранятся и сведения о состоянии оплаты услуг данным абонентом. Приписки АТ к тому или иному ЦК происходит при заключении с абонентом первоначального договора на обслуживание и остается неизменной, пока договор действует. При заключении такого договора в сотовый телефон абонента вставляется Subscriber Identity Module (SIM)-карта, содержащая, в частности, и идентификационный номер, под которым абонент будет далее фигурировать в домашнем или гостевых регистрах.

Процедура определения местоположения АТ выполняется самим терминалом, а ее результат сообщается на ЦК. Эта процедура сводится к тому, что АТ в интервалы времени, свободные от сеансов связи, поочередно настраиваются на прием сигналов с теми наборами параметров, которые могут в разных сотах использоваться для передачи ШРК. Те из них, на которых удастся принять такие сигналы (фиксируется при обнаружении соответствующей структуры сообщений), запоминаются АТ вместе с результатами измерения уровня данных сигналов на входе АТ. Принимается, что АТ находится в той соте, которой соответствует сигнал ШРК максимального уровня. Именно набор параметров этого сигнала и является адресом соты размещения данного АТ. Для того чтобы данные были своевременно переданы на ЦК, периодически, по командам этого центра, проводятся сеансы регистрации АТ, при которых все такие терминалы, находящиеся в рабочем состоянии, поочередно сообщают на ЦК адреса своих сот. Если в течение нескольких последовательных регистраций данный АТ не сообщил своего адреса, то считается, что этот терминал выключен или находится в недоступном для организации радиоканала месте. Полученные при регистрациях данные заносятся в регистры памяти ЦК, так что при получении запросов на организацию сеанса связи ЦК сразу может либо сообщить о недоступности данного АТ, либо послать вызов в определенную соту.

Если АТ перемещается достаточно быстро и размеры используемых сот не слишком велики, то время, в течение которого АТ может пересечь соту, оказывается соизмеримым с длительностью сеанса связи. В этом случае имеется заметная вероятность того, что за время сеанса связи АТ столь далеко отойдет от соты, в которой этот сеанс начинался, что может наступить нарушение связи. Для исключения такой ситуации в системе сотовой связи должен обеспечиваться режим передачи обслуживания из соты в соту в процессе сеанса связи. Реализация такого режима требует, по крайней мере, чтобы АТ в процессе сеанса связи мог продолжать следить за своим адресом, а также передавать его на ЦК. Последняя процедура требует наличия сопровождающего канала управления, а первая может быть реализована, только если АТ способен, кроме основного канала, одновременно принимать еще и ШРК. Пути реализации последнего требования без заметного усложнения приемного оборудования АТ описаны ниже, применительно к конкретным типам систем сотовой связи.

Выше уже упоминалась процедура роуминга. Под *роумингом* понимают возможность АТ, приписанного к сети некоторого географического района, работать в сети того же стандарта, но расположенной в другом районе. Роуминг может быть обеспечен при наличии каналов

связи между ЦК обоих регионов на основе соглашения между соответствующими операторами. При появлении некоторого АТ, приписанного к сети 1, в зоне действия ЦК сети 2, последней запрашивает необходимые данные из домашнего регистра первой сети, после чего обслуживает этот АТ, как и АТ своей сети. Так, в настоящее время установлен роуминг практически между всеми сетями GSM-900 России, а также между такими сетями и основными сетями этого стандарта европейских стран.

11.2.2. GSM: частотный план и многостанционный доступ

Для работы сотовых систем стандарта GSM в России и Западной Европе реализуются следующие диапазоны частот: 935–960 МГц для передачи в прямом направлении и 890–915 МГц в обратном. Каждый диапазон разделен на 124 стандартные частотные полосы (СЧП) по 200 кГц каждая. СЧП прямого и обратного направлений собраны в пары, так что сдвиг по частоте между ними составляет 45 МГц. Каждая такая пара дуплексных стандартных полос частот (ДСЧП) используется для организации дуплексных каналов связи между БС и АТ. ДСЧП присвоены номера от 1 до 124. Каждой соте выделяется несколько таких ДСЧП и при том таким образом, чтобы по крайней мере в соседних сотах ДСЧП с одинаковыми номерами не использовались. Распределение ДСЧП между сотами носит название *частотно-территориального планирования* и производится не оперативно эксплуатирующим персоналом на основании средних трафиков сот и закрепляется впредь до нового решения. В качестве варианта может устанавливаться расписание использования ДСЧП.

Стандарт GSM предусматривает применение сот радиусом до 35 км. Реальные размеры сот зависят от высоты размещения БС соты, застройки окружающей территории и целого ряда других факторов.

Отметим, что на основе стандарта GSM в настоящее время разработаны и внедрены определенные его модификации, отличающиеся, в частности, и диапазонами используемых частот. Так, в Европе и России используется стандарт DCS-1800 (Digital Cellular System), отличающийся сокращенным размером сот и диапазоном частот (1805–1880 МГц для передачи в прямом направлении и на 95 МГц ниже в обратном; всего здесь организуется 374 ДСЧП).

В настоящее время в нашей стране все чаще используют двумодовые сотовые телефоны, обеспечивающие возможность работы как в сети GSM-900, так и в сети DCS-1900. В Северной Америке используют вариант GSM, работающий в диапазоне 1900 МГц (стандарт PCS-1900).

Стандарт GSM предусматривает работу в каждой ДСЧП нескольких АТ на общей несущей с пакетным временным разделением. Длительность временного интервала (слота), отводимого для передачи каждого пакета, составляет 0,546 мс; 8 слотов (пакетов, вместе с защитными интервалами между ними) образуют кадр длительностью 4,615 мс. Скорость передачи бит внутри пакета составляет $D_m = 270,833$ кбит/с, так что время передачи одного бита в пакете — 3,69 мкс, а число бит в нем — 156,25. Из этого количества 114 бит представляют закодированное полезное сообщение, а 26 бит, размещающихся посередине пакета, содержат заранее известное на приемной стороне обучающее слово. Последнее используется для синхронизации (см. гл. 10) и настройки эквалайзера (см. гл. 9). Остальные биты пакета служат для служебных целей.

Для организации стандартного телефонного канала используется по одному пакету на кадре. Таким образом, если бы все слоты всех кадров выбирались только для передачи информации, скорость передачи бит, представляющих полезный кодированный сигнал, составила бы $D_k = 114 \text{ бит}/4,65 \text{ мс} = 24,516$ кбит/с. В действительности, часть слотов используется

для организации тех или иных каналов управления. Поэтому скорость передачи кодированных информационных бит оказывается меньше D_k . В соответствии с порядком использования кадров они объединяются в мультикадры, которые, в свою очередь, объединяются в суперкадры, а последние — в гиперкадры, длительность которых достигает уже нескольких часов. Использование гиперкадров определяется процедурами защиты информации, передаваемой по линиям сотовой связи.

Кроме полноскоростных телефонных каналов, стандарт GSM предусматривает и каналы половинной скорости, пакеты которых передаются не на каждом информационном кадре, а через один такой кадр. Наоборот, для повышения скорости передачи информации от АТ, что в настоящее время является актуальным для целей обеспечения таких новых услуг, как передача данных и работа в сети Интернет, в интересах данного АТ может использоваться несколько пакетов за кадр.

11.2.3. GSM: радиointерфейсы и характеристики оборудования

В системах GSM используется гауссовская модуляция с минимальным сдвигом (см. 6.2). При этом скорости передачи бит D_m соответствует скорость передачи ЭС $B = D_m/2 = 135,417$ кбит/с. Поскольку для передачи такого радиосигнала используется полоса 200 кГц, то удельная скорость передачи $\gamma = 0,68$ симв/с/Гц = 1,36 бит/с. Это соответствует значениям спектральной эффективности ММС, определенной в 6.2.

Для повышения помехоустойчивости стандарт предусматривает помехоустойчивое кодирование каскадным кодом с внешним блочным (53,50,3) и внутренним сверточным кодами. Последний имеет длину кодового ограничения 5 при скорости кода $1/2$ и декодировании в целом по алгоритму Витерби. С учетом избыточности кода и отмеченном в предыдущем пункте использовании части кадров для управления системой скорость передачи данных по полноскоростному каналу составляет 9,6 кбит/с.

Для того чтобы обеспечить независимость ошибок в кодовой комбинации и, тем самым, обеспечить эффективность кодирования в каналах с замираниями, стандарт предусматривает реализацию перемежения кодированных сообщений. Глубина перемежения определяется вызываемой им задержкой в передаче данных по радиолинии, которая принята равной величине порядка 2000 бит. Как отмечалось в гл. 9, глубина перемежения при кодировании должна быть тем большей, чем больше интервал временной корреляции замираний. Поскольку последняя величина случайна, нельзя гарантировать, что принятая глубина перемежения окажется достаточной. Для борьбы с этим недостатком стандарт GSM предусматривает использование псевдослучайной перестройки рабочей частоты (ППРЧ) в процессе передачи. Именно несущая частота или, что то же самое, номер ДСЧП, используемый находящимися на связи АТ, изменяется каждый кадр. При этом терминалы как бы взаимно обмениваются несущими частотами, так что реализация ППРЧ не требует увеличения общего числа рабочих частот. Изменение рабочей частоты должно, как правило, превосходить интервал частотной корреляции замираний. Тогда замирания сигнала на различных кадрах или на различных пакетах любого канала оказываются независимыми. Таким образом, ППРЧ приводит к ограничению временного интервала корреляции замираний длительностью одного пакета, что обеспечивает достаточность используемой глубины перемежения кодовых символов.

Наконец, стандарт GSM предусматривает еще метод борьбы с селективными замираниями, связанный с измерением текущей частотной характеристики канала связи и компенсацией возникающих ее отклонений. Речь идет об использовании эквалайзера, принцип действия которого был описан в гл. 9. Настройка эквалайзера предусматривается на основе анализа обучающего слова, наличие которого в каждом пакете было отмечено выше.

Мобильные телефоны стандарта GSM (ручные АТ) имеют небольшую массу (до 150 г), антенну, длина которой не превосходит 10 см и передатчик с пиковой мощностью до 0,8 Вт. Следует иметь в виду, что за счет временного разделения средняя мощность передатчика оказывается существенно меньшей пиковой. Так, при организации полноскоростного телефонного канала средняя мощность, по крайней мере, не превосходит 100 мВт.

Следует подчеркнуть, что стандарт предусматривает временное разделение между приемом и передачей на АТ. Принимаемые и передаваемые данным АТ пакеты не перекрываются во времени, так что прием и передача никогда не ведутся АТ одновременно. Сдвиг между приемом и передачей предусмотрен не менее 1,092 мс, т.е. двух слотов, что, по крайней мере, на порядок больше времени распространения сигнала по линии АТ–БС. Поэтому различие времен распространения для отдельных АТ, вызванное отличием расстояний до БС, может не учитываться при назначении слот для приема и передачи. Временное разделение между приемом и передачей значительно упрощает реализацию АТ, поскольку исключает необходимость в применении сложных фильтровых устройств развязки между приемником и передатчиком терминала.

Оборудование БС стандарта GSM, в зависимости от класса обеспечивают одновременное обслуживание от двух до шести дуплексных несущих. Предусматривается возможность наращивания пропускной способности за счет установки на БС нескольких комплектов такого оборудования.

11.2.4. CDMA. Общая характеристика

Стандарт CDMA (IS-95) предусматривает кодовое разделение каналов, в связи с чем предписывает использование широкополосных фазоманипулированных сигналов (ФМ ШПС, см. гл. 6 и 7). Как было отмечено, такие сигналы, наряду с известными достоинствами, обладают тем недостатком, что создают взаимные помехи, так что именно эти взаимные помехи ограничивают число каналов, которые могут одновременно функционировать в выделенной полосе рабочих частот. По этой причине стандарт предусматривает реализацию всех возможных мер, позволяющих сократить уровень взаимных помех между каналами. Соответствующие требования и определяют специфические особенности методов передачи и приема, используемых в сотовых CDMA системах связи. Охарактеризуем эти особенности.

Ясно, что помехи, создаваемые сигналом данного канала остальным сигналом тем больше, чем выше мощность сигнала данного канала. Номинальная мощность АТ выбирается так, чтобы соответствовать самым неблагоприятным из возможных условий связи. Поэтому в большинстве случаев она оказывается избыточной. Стандарт предусматривает регулировку мощности радиосигналов как в прямом (от БС на АТ), так и в обратном (АТ–БС) направлениях в процессе каждого сеанса связи в зависимости от текущей надежности данного радиоканала, причем точность и скорость слежения по уровню жестко нормируются.

Другая возможность сокращения взаимных помех заключается в выборе для отдельных каналов широкополосных сигналов, попарная ортогональность которых выполняется как можно более строго. Эта цель может быть достигнута, если отдельные сигналы синхронизованы во времени по такту и циклу используемых ПСП. Стандарт предусматривает соответствующую синхронизацию в тех случаях, когда ее осуществление не связано с чрезмерными техническими трудностями. Именно такая синхронизация предусматривается для сигналов, излучаемых БС в прямом направлении и не предусматривается для сигналов в обратных каналах. Дело в том, что АТ разнесены по пространству так, что различие времен распространения их сигналов до БС часто значительно превосходит требуемую

точность синхронизации. Именно по этой причине обеспечение их синхронного прихода на вход приемника БС представляет сравнительно сложную задачу.

В тех случаях, когда синхронизация по такту и циклу достигается, важно так выбрать используемые для сигналов различных каналов ПСП, чтобы они были ортогональны как можно более строго. Стандарт IS-95 предусматривает поэтому использование в прямом направлении в качестве таких ПСП функций Уолша (см. 4.3), которые теоретически строго ортогональны.

Еще одна особенность используемых в CDMA системах методов передачи и приема заключается в полезном использовании пауз, которые могут возникать при формировании сообщений абонентами. Так, при дуплексной телефонной связи оба абонента, как правило, не говорят одновременно, так что уже по этой причине возникают паузы, занимающие в среднем 50 % общего времени передачи в каждом направлении. Далее в речи неизбежны интонационные и другие паузы, так что в каждом направлении канал связи оказывается действительно занятым не более 30% времени. Аналогичная ситуация, как правило, имеет место и при факсимильной передаче, или передаче данных; такие данные передаются только в одну сторону в то время как обратный канал, используемый лишь для передачи служебных сигналов, оказывается почти не загруженным. Если во время пауз существенно снижать мощность радиосигнала, то можно заметно сократить уровень взаимных помех системы.

Исследования показали, что в системах телефонной связи целесообразно различать не только речь и паузы, но и учитывать несколько возможных градаций устойчивости элементов речи по отношению к помехам (показатели речевой активности). Так, громкая речь оказывается более устойчивой, чем тихая, а гласные звуки более устойчивы, чем согласные. Соответственно при преобразовании речи в цифровую форму может одновременно вырабатываться и показатель речевой активности (ПРА). Стандарт IS-95 предусматривает изменение на основании ПРА скорости передачи информации и соответственно мощности радиосигнала. При минимальной речевой активности, соответствующей паузе речи, скорость передачи информации $C_0 = 1,2$ кбит/с. Следующий уровень ПРА соответствует скорости $C_1 = 2,4$ кбит/с, следующий $C_2 = 4,8$ кбит/с, а при максимальном значении ПРА скорость составляет $C_3 = 9,6$ кбит/с.

Пропорционально скорости передачи информации изменяется и мощность сигнала. Указанные изменения скорости передачи информации и соответственно мощности сигнала происходят в темпе изменений динамики речи и являются сравнительно быстрыми. Что же касается изменений мощности сигнала под действием регулировки в зависимости от текущей надежности канала (см. выше), то они являются существенно более медленными, причем регулировке подвергается среднее (по ПРА) значение мощности.

Изменение скорости передачи информации происходит без изменения скорости передачи бит D за счет соответствующего изменения избыточности используемого помехоустойчивого кода. Сам код включает две ступени. Вначале сверточный код с длиной кодового ограничения $k = 9$ и скоростями: $R = 1/2$ для прямого и $R = 1/3$ для обратного каналов, а потом простейший код, сводящийся к повторению символов. Число повторений выбирается таким, чтобы для прямого направления всегда $D = 19,2$ кбит/с, а для обратного $D = 28,8$ кбит/с. Соответственно, для скорости C_3 повторения не используются, для C_2 каждый кодовый символ повторяется 2 раза, для C_1 — 4 раза, для C_0 — 8 раз.

С целью борьбы с замираниями стандарт предусматривает перемежение кодовых бит, осуществляемое путем использования матрицы памяти с различным порядком записи и считывания кодовых бит. Этот порядок однозначно установлен в стандарте для каждой скорости передачи как для прямого, так и для обратного каналов. В прямом канале при этом используется матрица памяти с 24 строками и 16 столбцами (всего 384 ячейки памяти), а в обратном — с 32 строками и 18 столбцами (576 ячеек памяти). Отметим, что время,

в течение которого заполняется матрица, составляет 20 мс как для прямого, так и для обратного каналов (с учетом отличия их битовых скоростей D). Этот интервал именуется *фреймом* и служит единицей отсчета системного времени.

Остановимся теперь на методах модуляции и демодуляции, предусмотренных стандартом IS-95. В прямом канале при передаче кодовой последовательности используется двоичная фазовая манипуляция при когерентном приеме и декодировании в целом по алгоритму Витерби. В обратном канале предусмотрена 64-ичная манипуляция, использующая 64 ортогональных радиосигнала. В качестве таковых применяются сигналы, образованные в результате двоичной фазовой манипуляции несущей функциями Уолша, длиной в 64 двоичных символа. Демодуляция в обратном направлении некогерентная, а декодирование осуществляется также по алгоритму Витерби.

11.2.5. CDMA: формирование широкополосных сигналов

Стандарт предусматривает использование ШПС с тактовой частотой дополнительной (расширяющей спектр) манипуляции 1,2288 МГц (длительность элементарного символа $\tau_0 = 0,814$ мкс). Как уже отмечалось выше, в прямом направлении все излучаемые БС сигналы синхронизируются по такту и циклу расширяющих спектр последовательностей (ПСП). Для этого на всех БС используется единое время, получаемое от навигационных спутниковых систем, например от GPS. Часы единого времени CDMA систем были запущены одновременно с часами GPS 6 января 1980 г. в 00.00.00. При формировании сигналов стандарт IS-95 предусматривает использование в качестве ПСП длинных и коротких M -последовательностей и функций Уолша. Длинные последовательности имеют период $T_d = 2^{41}\tau_0$, что составляет около 466 ч. Период коротких последовательностей составляет $T_k = 2^{15}\tau_0 = 2/75$ с. Указанные последовательности формируются с использованием M -последовательностей максимальной длины с n , равными 41 и 15, соответственно.

За начало периода таких последовательностей принимается момент появления первой единицы после 40 или 14 нулей соответственно (период заканчивается 40 или 14 нулями подряд). Используемые последовательности отличаются от таких M -последовательностей добавлением одного нуля в конце каждого периода, который, таким образом, оканчивается уже 41 или 15 нулями. Моменты начала формируемых на БС длинных и коротких ПСП синхронизируются с системным временем: именно с указанным выше моментом запуска системного времени совмещается начало первых периодов длинной ПСП, а также некоторых опорных коротких ПСП. Расчеты показывают, что следующий раз моменты начала периодов этих ПСП совместятся не ранее чем через 37 веков!

Блок-схема устройства формирования ФМ ШПС, транслируемого в прямом направлении, представлена на рис. 11.1, *а*.

На рисунке ГДПС — генератор длинной последовательности, вид которой определяется ключом того абонента, которому предназначен формируемый радиосигнал. В прямом направлении длинная ПСП (ДПС) используется только для шифрации сообщения. Для этого ДПС подвергается процедуре децимации в 64 раза (в блоке ДЦ-64), т.е. из этой последовательности выделяется только каждый 64 бит. Соответственно, частота следования бит сокращается в 64 раза, достигая величины $1,2288 \text{ МГц}/64 = 19,2 \text{ кГц}$, так что эти биты становятся синхронными с кодовой последовательностью на выходе перемежителя. Для шифрации две последовательности суммируются по mod 2 (такое суммирование на рис. 11,1, *а* обозначено крестом внутри кружка). Для дешифрации в месте приема последовательность демодулированных бит должна быть просуммирована по mod 2 с той же прореженной ДПС.

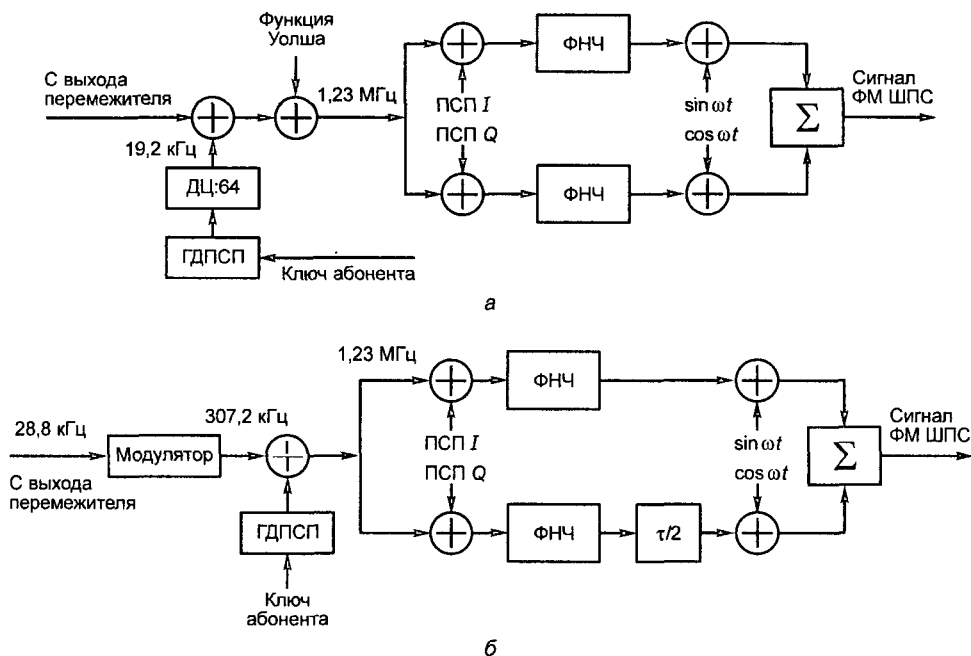


Рис. 11.1. Формирование ФМ ШПС в прямом (а) и обратном (б) каналах

Опишем теперь дополнительную манипуляцию, расширяющую спектр сигнала. В качестве основного адреса, определяющего АТ, которому предназначен формируемый сигнал, используется одна из 64 ортогональных функций Уолша с периодом 64 двоичных символа и тактовой частотой 1,2288 МГц. Для создания более равномерного спектра ШПС реализуется еще манипуляция короткими ПСП (КПСП). Одновременно с помощью этих кодов передается адрес БС, позволяющий АТ отличать одну БС от другой. Формируются две КПСП: синфазная ПСП и квадратурная ПСП. Тактовая частота этих последовательностей составляет 1,2288 МГц, а их вид зафиксирован в стандарте. Адрес БС определяется сдвигом начала периодов данных последовательностей по отношению к системному времени. Это смещение характеризуется индексом смещения l , принимающим 512 значений (от 0 до 511). Величина смещения, выраженная в битах КПСП, равна $n = 64l$. Смещение может отсчитываться от любой четной секунды системного времени, поскольку в двух секундах укладывается целое число периодов КПСП (75 периодов).

Формирование ШПС в обратном канале заметно отличается от описанного (рис. 11.1, б).

Здесь прошедшие переключатель кодовые комбинации поступают на модулятор, ставящий в соответствии блоку из шести кодовых бит одну из 64-разрядных ортогональных функций Уолша. Тем самым осуществляется шестикратная (64-ичная) модуляция, использующая ортогональные сигналы. Легко понять, что при такой манипуляции битовая скорость передачи D возрастает в 64/6 раз. (На этом и предыдущем рисунках рядом с теми или иными точками схемы указаны тактовые частоты цифровых сигналов, присутствующих в этих точках.) Далее спектр сигнала расширяется за счет наложения ДПСП. Последовательности различны для различных АТ, поскольку определяется индивидуальным ключом данного абонента. Накладываемые на сигнал обратных каналов в квадратурах I и Q КПСП

однозначно зафиксированы в стандарте и одинаковы для всех АТ системы. При наложении КПСП используется двукратная фазовая манипуляция со сдвигом. Этот сдвиг создается задержкой на $t/2 = 407$ нс.

11.2.6. CDMA: частотный план и параметры аппаратуры

Работа системы CDMA предусматривается в диапазоне частот 870–895 МГц (прямое направление) и 825–850 МГц (обратное). Сдвиг частот 45 МГц. Границы диапазона указаны ориентировочно, поскольку стандартом предусматривается несколько различных вариантов использования частот. Что касается шага сетки частот, то и АТ, и БС должны предусматривать возможность настройки на любую частоту диапазона с шагом 30 кГц. Вместе с тем в каждом конкретном случае используются частоты с шагом 1260 кГц, причем на каждой такой частоте с одной БС могут одновременно передаваться сигналы до 64 каналов (по числу функций Уолша, используемых для адресации каналов в прямом направлении). Именно выделение данного АТ той или иной функции Уолша и определяет предоставление ей канала связи в системе. По завершении сеанса связи с этой АТ данная функция Уолша освобождается и может быть предоставлена другому абоненту.

Что касается рекомендуемых стандартом параметров аппаратуры, то предусматриваются три класса АТ. Станции первого класса должны, при подаче команды на работу с максимальной мощностью обеспечивать на входе передающей антенны ВЧ мощность от 0,5 до 2,5 Вт, второго — от 0,2 до 1 Вт, а третьего — от 0,08 до 0,4 Вт.

Для того чтобы разделять отдельные лучи, возникающие при многолучевом приеме (см. гл. 9), в приемниках как БС, так и АТ предусматривается реализация нескольких независимых трактов приема, каждый со своей системой синхронизации по задержке. В каждой из этих станций в соответствии со стандартом должно быть не менее четырех таких трактов для каждого канала. Один из них используется для панорамного просмотра всех принимаемых лучей и оценки их текущих уровней. В результате отбирается не менее трех наиболее сильных лучей, которые постоянно принимаются остальными трактами приема и после свертки по частоте когерентно суммируются. Образованный суммарный сигнал демодулируется, демультиплексируется и декодируется для получения принимаемого сообщения. Если при панорамном просмотре выясняется, что один из неиспользуемых лучей возрос по уровню и превзошел некий используемый луч, то производится соответствующее изменение состава используемых лучей. Поскольку при этом, как правило, заменяется только один луч, то соответствующее изменение проходит незаметно для пользователя. Алгоритм может быть использован также при смене БС во время сеанса связи. Это может стать необходимым в связи с перемещением АТ из одной соты в другую. При этом АТ одновременно принимает сигналы от двух БС, один из которых нарастает, а другой убывает; суммирование сигналов и позволяет плавно перейти от приема БС одной соты («старой») к БС другой («новой») соты.

Такой «мягкий» переход из соты в соту учитывается и при формировании самих сот. При CDMA формирование сот осуществляется следующим образом: вначале вся зона обслуживания делится на области в виде кругов (гексагональная сетка, см. 6.1.4), а затем каждая такая область делится на три радиальных соты. Обслуживающие их три БС размещаются в центре области. На БС используют антенны с шириной диаграмм направленности в азимутальной плоскости 60° . Диаграммы направленности этих БС представлены, сплошными линиями на рис. 11.2 (штриховыми линиями показаны границы трех радиальных сот).

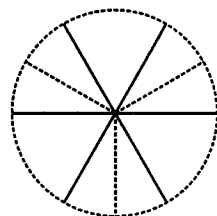


Рис. 11.2. Структура сот при CDMA

Если АТ находится в створе диаграммы направленности своей соты, то он работает с одной БС (своей). Если он оказывается вне створа этой диаграммы (вблизи границы между сотами), то он обслуживается двумя БС (своей соты и другой, ближайшей к месту расположения АТ). Через обе эти БС для данного АТ передаются идентичные сигналы и они когерентно суммируются приемными трактами АТ. Такой выбор параметров антенн базовых станций энергетически оказывается более эффективным, чем очевидный вариант с использованием антенн с шириной диаграммы направленности по азимуту 120° при работе всегда только через одну БС.

11.3. Сотовые спутниковые системы подвижной связи

11.3.1. Принципы построения и показатели эффективности

В сотовых системах спутниковой связи заданная для обслуживания область поверхности Земли разбивается на соты, и каждая из них обслуживается спутником отдельно. Сотовое обслуживание (СО) может быть реализовано либо только на линиях вверх, либо только на линиях вниз, либо на линиях обеих этих групп. Конкретно, СО на линиях вверх организуется следующим образом: подлежащий обслуживанию регион разбивается на некоторое число N сот, и на борту используется для каждой такой соты своя приемная антенна, диаграмма направленности которой выбрана такой, чтобы освещать только данную соту, не затрагивая, по возможности, остальные. Следует различать параллельное и последовательное СО. В первом случае на борту должно существовать N приемных антенн, и прием сигналов из всех сот должен вестись параллельно. Во втором — может использоваться только одна антенна, которая поочередно (циклически) освещает отдельные соты. Соответственно, ЗС каждой соты собирают информацию, подлежащую передаче по линии вверх, в пакеты и передают их в том кадре цикла перенацеливания бортовой приемной антенны, когда освещается именно эта сота.

Аналогично и по передаче можно рассматривать параллельное и последовательное СО. В первом случае на борту создается N независимых трактов передачи и сигналы во все соты передаются параллельно, а во втором — используется один тракт передачи, антенна которого циклически освещает отдельные соты. Хотя при параллельном обслуживании антенны должны функционировать параллельно, это не значит что они не могут иметь общих элементов. Такие антенны создаются, например, с использованием общего зеркала, облучаемого матрицей облучателей. Если центральный облучатель матрицы находится в фокусе зеркала, то остальные оказываются в той или иной степени вынесенными из него, так что создаваемые ими лучи диаграммы направленности антенны соответственно отклоняются от оси зеркала. Таким образом, с помощью одного зеркала формируется веер лучей, причем каждый из них имеет свой независимый выход (для приемных антенн) или вход (для передающих). Таким образом, создаются многолучевые антенны (МЛА), каждый луч которых освещает свою соту.

Остановимся на оценке энергетической и полосной эффективности СО в спутниковой связи при передаче индивидуальной информации. Природа соответствующих выигрышей принципиально та же, что и в наземной связи (см. 6.1.4). Однако описывающие их количественные соотношения отличаются. Остановимся на них специально.

Начнем с рассмотрения энергетической эффективности для случая реализации КА на ГСО, когда для расчета спутниковых радиолиний могут использоваться соотношения параграфа 11.1. За базу для сравнения примем вариант, когда на борту СР по приему и передаче

используется антенна, освещающая сразу всю зону обслуживания. В отличие от сотового, такой вариант будем называть *интегральным обслуживанием* (ИО). Пусть при ИО коэффициент усиления бортовой приемной антенны составляет $G_{\text{он}}$. Тогда при переходе на СО он для каждого луча в идеале должен увеличиться в N раз и составить $G_{\text{ос}} = NG_{\text{он}}$. Это означает, что и приведенная добротность приема на СР возрастет в N раз. Если средняя скорость передачи информации от ЗС по линии вверх не изменится, то для обеспечения того же соотношения сигнал/шум на входе борта, который имел место при ИО, потребуется средняя мощность передатчика ЗС в N раз меньше, чем при ИО.

Если по приему на СР реализуется параллельное СО, то и пиковую мощность передатчика ЗС можно будет во столько же раз снизить. При последовательном СО пиковая мощность данного передатчика должна быть сохранена, однако эту мощность передатчик должен излучать лишь $1/N$ общего времени работы (на одном кадре за цикл перенацеливания антенны); на остальных кадрах передатчик может отключаться.

Таким образом, во всех случаях переход от ИО к СО на линиях вверх приводит к снижению требуемой средней мощности ЗС и, следовательно, к существенному повышению энергетической эффективности таких линий. Поскольку соответствующий выигрыш по мощности имеет порядок числа сот, а оно практически может достигать нескольких сотен, то это позволяет резко ослабить требования к ЗС по ЭИИМ.

Аналогично можно охарактеризовать энергетический выигрыш от СО на линии вниз. При интегральном обслуживании по линии вниз должна передаваться сразу вся информация, предназначенная для приема на всех ЗС системы. При СО в каждый луч должна передаваться только та часть информации, которая предназначена ЗС данной соты. Поэтому в среднем в каждый луч передается N -я часть общего информационного потока. Далее, и коэффициент усиления луча передающей антенны возрастает при СО в N раз. Эти два обстоятельства обеспечивают вместе энергетический выигрыш в N^2 раз. Однако, если еще учесть что суммарная ВЧ мощность СР должна быть при СО распределена по N лучам, то найдем, что результирующий выигрыш составит N раз. Легко видеть что этот результат имеет место как при параллельном, так и при последовательном СО. Вышеизложенное показывает, что это утверждение справедливо как для приема на борту, так и для передачи с СР.

Остановимся на полосной эффективности, обеспечиваемой при СО. Основой для совершенствования соответствующих характеристик служит повторное использование частот, которое достигается за счет пространственной селекции. И здесь может быть использован семичастотный или девятичастотный план повторного использования частот, что повышает спектральную эффективность в $(0,11 - 0,15)N$ раз. Правда это относится только к параллельному СО. При последовательном СО все лучи могут работать в одной и той же полосе частот, поскольку сигналы отдельных лучей все равно разделены по времени. При этом занимаемая системой полоса частот сохраняется такой же, как и при интегральном обслуживании (поскольку сохраняется и общий объем транслируемой информации), так что выигрыш в полосной эффективности отсутствует.

Изложенное позволяет утверждать, что в спутниковой, как и в наземной радиосвязи именно использование СО является основным путем повышения энергетической и полосной эффективности. Тот факт, что в настоящее время такое обслуживание используется только в подвижной связи, определяется тем, что именно в подвижной связи наблюдается особый дефицит как ВЧ-мощности, так и полосы рабочих частот. При этом дефицит ВЧ-мощности определяется тяжелыми условиями распространения радиоволн (затенения, многолучевость и т.п.), ограниченностью источников питания и экологическими ограничениями, а жесткая лимитированность полосы связана с использованием сравнительно низкочастотных диапазонов волн, диктуемым требованием сравнительно широких диаграмм направленности антенн.

11.3.2. Глобальные и региональные системы

Для создания спутниковых систем подвижной службы особенно удобно использовать СР, размещаемые на сравнительно низких орбитах (высота над Землей порядка тысячи км). Число спутников, необходимое для полного покрытия зоны обслуживания, приближающейся к глобальной, составляет 50–100. Если даже число сот обслуживаемых с каждого спутника составляет несколько десятков, то общее число сот приходящихся на всю зону обслуживания (создаваемых всеми СР системы) имеет порядок $N = 1000$ и более, что в соответствии с вышеизложенным, обеспечивает очень большое, кардинальное повышение энергетической и полосной эффективности системы связи. В конце прошлого века были созданы и введены в эксплуатацию две такие системы подвижной связи: «Globalstar» и «Iridium». Краткое описание этих систем приведено ниже.

Эффективное использование глобальных систем подвижной спутниковой связи требует, в конечном счете, чтобы почти на всей территории нашей планеты существовал достаточно высокий платежеспособный спрос на соответствующие услуги. Как показал опыт, это требование пока не реализуется в достаточной мере, так что рост числа пользователей указанных систем происходил гораздо медленнее, чем ожидалось, в связи с чем их владельцы начали испытывать значительные финансовые трудности. Это привело к фактическому прекращению коммерческой эксплуатации системы «Iridium», которая являлась технически особенно совершенной и, потому, дорогой. Указанные трудности привели к повышению интереса к региональным системам подвижной спутниковой связи, ибо можно ожидать, что рациональный выбор зоны обслуживания обеспечит достаточный платежеспособный спрос на услуги системы.

К сожалению, создание региональных систем спутниковой связи с использованием низколетящих СР оказывается гораздо менее выгодным. Дело в том, что по мере сокращения зоны обслуживания число требуемых низколетящих СР практически не уменьшается. Все равно, необходимо обеспечивать непрерывное покрытие заданной зоны, а применительно к низколетящим спутникам с периодом обращения порядка двух часов это означает и непрерывное покрытие практически любых других зон. По этой причине часто оказывается, что экономически целесообразнее создавать региональные системы с использованием СР, размещаемого на ГСО. В этом случае достаточно одного спутника. Правда, для сокращения размеров сот приходится при этом использовать диаграммы направленности антенн с очень узкими лучами, что соответствует размещению на СР антенн большой площади. Тем не менее размеры сот получаются сравнительно большими, так что и требуемые мощности излучений в каждый луч оказываются весьма значительными. Если теперь учесть еще, что сокращение ширины лучей приводит к увеличению их количества, то станет ясно, что речь здесь может идти об использовании в качестве СР только тяжелого космического аппарата с достаточно мощными источниками электропитания. В последние годы были введены в эксплуатацию две такие системы подвижной спутниковой связи: система, принадлежащая ОАЭ со спутником «Thuraya» и система ACeS (Asia Cellular Satellite) со спутником «Garuda», созданная ассоциацией стран Юго-Восточной Азии. Идеологически эти системы достаточно близки, хотя и реализуют различные стандарты сотовой спутниковой связи (первая GMR (Geo Mobile Radio)-1, а вторая GMR-2). Более удачной представляется первая из этих систем; на ее описании мы остановимся ниже.

11.3.3. Система «Globalstar»

Система сданная в эксплуатацию в 1999 г. использует 48 одновременно действующих СР, размещенных на шести круговых орбитах с высотой 1414 км над поверхностью Земли. Плоскости орбит наклонены к плоскости экватора под углом 55° , а прямые пересечения

плоскостей орбит с экваториальной плоскостью образуют веер с шагом по углу 30° . В каждой плоскости вращаются 8 СР (интервал по углу между ними 45°). Такая орбитальная группировка обеспечивает возможность непрерывного обслуживания всей территории Земли, исключая заполярные области (северную и южную). На большей части обслуживаемой территории, особенно по мере приближения к экватору, с большой вероятностью в зоне видимости абонентов находится сразу несколько спутников. Задача системы «Globalstar» сводится к обеспечению соединения подвижных абонентских терминалов (АТ) со стационарными пунктами привязки каналов этой системы к наземным сетям общего пользования. В этих пунктах размещаются стационарные шлюзовые станции спутниковой связи (ШС), через которые такая привязка и обеспечивается. Количество и размещение ШС должны быть такими, чтобы любой СР системы находился в зоне видимости хотя бы одной такой станции. Практически число таких станций превышает минимально требуемое их количество, поскольку каждая страна, использующая эту систему связи, предпочитает, обычно, иметь у себя обслуживаемую ее территорию ШС. В соответствии с задачами системы, каждый ее СР обеспечивает трансляцию сигналов между АТ и ШС, находящимися в поле его зрения. По мере полета спутника меняется видимый с него регион Земли, а следовательно, и работающие через него АТ и ШС. Регион же, который обслуживался им ранее, обслуживается теперь другими СР системы.

Отметим, что в системе «Globalstar» при необходимости обеспечения связи между АТ, используется двухскачковая линия, а именно: первый АТ через СР, обслуживающий в данное время область его размещения, связывается с ШС, работающей с этим СР, которая по наземной системе телефонной связи транслирует соответствующие сигналы к той ШС, которая может находиться на связи со вторым АТ. Последняя, через СР, обслуживающий второй АТ, транслирует на него эти сигналы.

Космические аппараты системы «Globalstar» имеют массу порядка 530 кг и солнечные батареи с мощностью электропитания 1200 Вт. Полезная нагрузка каждого КА имеет массу 152 кг и обеспечивает работу в трех диапазонах частот:

- по передаче на АТ S-диапазон (2483,5–2500 МГц);
- по приему от АТ L-диапазон (1610–1626,5 МГц);
- фидерная линия связи с ШС С-диапазон (линия вниз 6875–7075 МГц, линия вверх 5091–5250 МГц).

На линиях связи с АТ используется параллельное СО с разделением зоны видимости с каждого СР на 16 сот. Соты, формируемые одним СР, могут быть разбиты на три группы:

- центральная сота, с центром в подспутниковой точке;
- средние соты (6 штук), образующих первое кольцо вокруг центральной соты;
- крайние соты (9 штук), образующие второе кольцо.

Для формирования сот на борту используются многолучевые активные фазированные антенные решетки (АФАР) с 61 элементом по приему и 91 элементом по передаче. Соответственно каждый элемент приемной решетки снабжен МШУ, а каждый элемент передающей — твердотельным усилителем мощности. Последний обеспечивает максимальную ВЧ-мощность, равную 4,2 Вт, так что суммарная мощность излучения на передачу с борта в S-диапазоне может достигать 380 Вт. Каждая АФАР имеет столько выходов (входов) сколько лучей создается. Распределение ВЧ-мощности передающей АФАР между лучами определяется уровнями возбуждения соответствующих входов, т.е. числом каналов принимаемых данным лучем. Таким образом обеспечивается автоматическое перераспределение суммарной ВЧ-мощности между лучами в соответствии с текущим трафиком.

Диаграммы направленности лучей антенн выбраны так, чтобы компенсировать различные расстояния от СР до поверхности Земли в зависимости от направления связи. Такие

диаграммы направленности лучей именуется *равносигнальными*. Дело в том что дальность связи увеличивается по мере отклонений направления связи от вертикальной (по отношению к Земле) оси; для СР «Globalstar» это расстояние изменяется от 1400 км (при отсутствии отклонения) до 4400 км (в направлении по касательной к поверхности Земли), т.е. грубо говоря в три раза. Для компенсации этого различия необходимо обеспечить возрастание коэффициента усиления лучей по мере отклонения обслуживаемых ими сот от центральной. Именно, размеры сот должны быть выбраны так, чтобы телесные углы видимости со спутника крайних сот были бы раз в 5–8 (квадрат отношения средних расстояний) меньше, чем центральной соты. В действительности для СР «Globalstar» усиление в максимуме диаграммы луча обслуживающего центральную соту составляет около 13 дБ, а крайние соты — порядка 18 дБ. Отметим, что использование равносигнальных диаграмм направленности бортовых антенн обеспечивает возможность использования одинаковых АТ во всей области обслуживания. В противном случае пришлось бы использовать АТ с различными энергетическими характеристиками в различных сотах, в зависимости от положения соты по отношению к используемому в данный момент для ее обслуживания СР.

Что касается используемых способов передачи информации (АЦП речевых сообщений, кодирование, модуляция, многостанционный доступ и т.п.), то в системе «Globalstar» реализуется стандарт IS-95, описанный в предыдущем параграфе. Соответственно, полоса частот, занятая ШПС сигналом, составляет порядка 1,23 МГц, так что в выделенных для работы системы *L* и *S* диапазонах, в интервалах частот шириной 16,5 МГц в каждом, размещается 13 таких полос.

Частотный план полезной нагрузки СР «Globalstar» представляется достаточно своеобразным. Именно, он построен так, чтобы элементы ретранслятора не определяли бы распределение полосы рабочих частот между сотами (лучами); каждый луч обеспечивает ретрансляцию сигнала на любой из частот, принадлежащих полосе шириной 16,5 МГц. Исключение взаимовлияния сигналов отдельных лучей должно обеспечиваться при этом рациональным назначением рабочих частот для каждого организуемого канала связи, так чтобы эти частоты не повторялись в слишком близких сотах. За такую «свободу» в применении частот в *L* и *S*-диапазонах приходится платить расширением полосы используемых частот фидерной линии. Именно, в *S*-диапазоне дважды (за счет поляризационной развязки) используются полосы частот шириной порядка 160 МГц. В них создаются 16 приемных и 16 передающих трактов с полосой каждый порядка 20 МГц (по 8 трактов на каждой из двух ортогональных круговых поляризаций). Каждый луч имеет свой такой тракт, так что на борту создается 16 независимых дуплексных трактов ретрансляции сигналов между АТ и ШС (по одному для каждой соты) с полосой пропускания каждого 16,5 МГц.

Это, однако, не означает, что вся полоса может полностью использоваться в каждом луче при их одновременной работе. Действительно, как отмечалось при описании стандарта IS-95, на каждой несущей может быть создано до 64 каналов связи за счет использования различных расширяющих спектр ПСП. Все эти каналы могут создаваться как в интересах одной соты, так и в интересах различных, в том числе и близко расположенных сот. Однако каналы на той же несущей и с той же ПСП могут повторяться только в наиболее удаленных сотах. Таким образом, для обеспечения эффективной работы системы необходимо при предоставлении каналов абонентам постоянно следить за распределением связных ресурсов (несущих частот и расширяющих спектр ПСП), так чтобы избежать роста взаимных помех сверх допустимых уровней. Выбранная структура ПН позволяет назначать частоты и изменять в широких пределах пропускные способности отдельных лучей без подачи каких-либо команд на борт, только путем управления наземными средствами системы. Этому способствует и то обстоятельство, что бортовая ВЧ мощность автоматически перераспределяется

между лучами в соответствии с текущим трафиком (см. выше). Тем не менее задача управления связными ресурсами в системе «Globalstar» при предоставлении каналов абонентам оказывается достаточно сложной, поскольку оценка взаимных помех в системе должна учитывать не только сигналы, транслируемые данным, но и всеми соседними спутниками, находящимися в зоне видимости АТ. Следует также учитывать, что сигнал передаваемый любым АТ ретранслируется всеми СР, находящимися в зоне видимости этого терминала, поскольку антенны терминалов практически близки к ненаправленным. Часто сигналы, проходящие на данную ШС через различные СР, могут быть когерентно просуммированы и, следовательно, полезно использованы. В ряде случаев, например когда речь идет о СР, находящихся одновременно в зоне видимости АТ, но не в зоне видимости ШС, сигналы, ретранслируемые другими СР, признаются помехами.

Что касается земного комплекса системы, то по реализуемым процедурам формирования и обработки сигналов он подобен соответствующему оборудованию наземной системы сотовой связи с CDMA. АТ персональной связи по своим габаритно-массовым характеристикам близки к «трубкам» наземной системы (масса порядка 200 г). При этом в них, как и в наземной системе, реализуется не менее четырех приемников для независимой свертки и когерентного сложения наиболее сильных из принимаемых лучей сигнала данного канала.

Типовой шлюз системы «Globalstar» снабжен четырьмя приемо-передающими антеннами диаметром порядка 5,5 м и работает в автоматическом режиме (без участия эксплуатационного персонала). ШС обеспечивает одновременную работу максимально с четырьмя, находящимися в поле его зрения СР орбитальной группировки системы.

11.3.4. Система «Iridium»

Система была введена в эксплуатацию в 1998 г. Однако, поскольку количество пользователей возросло недостаточно быстро, система стала испытывать финансовые трудности и к 2001 г. обанкротилась. Вначале рассматривался вариант ее ликвидации, но потом у нее появились новые владельцы и эксплуатация системы была продолжена, однако в гораздо более скромных масштабах. В частности, вместо ряда региональных шлюзов в системе в настоящее время функционирует только одна ШС, находящаяся на территории США. В связи с отсутствием Российской ШС, эксплуатация системы на территории нашей страны в настоящее время не разрешена.

Несмотря на столь неудачную сложившуюся историю практического использования, система «Iridium» является технически одной из самых совершенных систем спутниковой связи. В частности, она представляет собой единственную систему, способную обеспечить связь между станциями, расположенными в любых двух точках на поверхности Земли, без использования ретрансляции через какие-либо другие земные станции, или, тем более, без использования каких-либо наземных систем связи. Эта возможность достигается за счет реализации линий непосредственной связи между СР орбитальной группировки системы.

Орбитальная группировка системы «Iridium» состоит из 66 СР, размещенных на круговых орбитах, лежащих в шести плоскостях (по 11 СР в каждой плоскости). В идеале все эти плоскости пересекаются по оси вращения Земли, так что все орбиты можно считать полярными. В действительности плоскости орбит несколько отклонены от строго полярных, для того чтобы исключить высокую вероятность столкновения спутников над полюсами Земли. Положения СР на соседних орбитах смещены «по фазе», так что общий вид орбитальной группировки в любой момент времени обнаруживает как бы шахматный порядок размещения СР. Высоты орбит над поверхностью Земли составляют 780 км. Использование квази-

полярных орбит обеспечивает полное, без каких-либо исключений, обслуживание всей территории Земли.

Масса каждого СР составляет 700 кг, причем обеспечивается трехосная стабилизация КА на орбите. Энерговооруженность спутника составляет около 1400 Вт, из которых 700 Вт предоставляется для питания полезной нагрузки. На каждом СР организуются следующие радиолинии:

- линии связи с подвижными объектами вверх и вниз (ЛПВв и ЛПВн соответственно) в L -диапазоне волн (1616–1625,5 МГц);
- фидерные линии связи с ШС вверх и вниз (ЛШВв и ЛШВн); диапазон Ка, ЛШВв 29,1–29,3 ГГц, ЛШВн 19,4–19,6 ГГц;
- линии межспутниковой связи (ЛМС) по четыре дуплексным направлениям от каждого СР) в диапазоне Ка 23,18–23,38 ГГц.

По поводу этих линий необходимо отметить следующее: ЛПВв и ЛПВн организуются в одном и том же диапазоне частот. Разделения сигналов на этих линиях достигается методом временного деления: СР при связи с подвижными объектами попеременно работает на прием и передачу. Аналогичное решение принято относительно ЛМС, в которых также предусматривается попеременная работа на прием и передачу. Интересно отметить, что это решение позволяет сделать все СР орбитальной группировки идентичными. Действительно, если бы частоты и передачи и приема здесь отличались бы, то поскольку для ЛМС частота передачи одного спутника должна быть частотой приема для другого, то пришлось бы использовать СР нескольких модификаций.

Для организации ЛПВв и ЛПВн на СР используется шесть приемо-передающих антенн, причем каждая антенна обеспечивает обслуживание последовательное СО восемь сот. Таким образом, один СР обслуживает 48 сот. Каждая антенна представляет собой ФАР, причем формируемый луч последовательно циклически освещает обслуживаемые соты. Все шесть антенн работают одновременно, так что в целом каждый СР осуществляет последовательно-параллельное обслуживание: организуется шесть групп по восемь сот в каждой, причем группы обслуживаются параллельно, а соты в группах — последовательно. Отметим еще, что поскольку используются невысокие орбиты, угол видимости Земли со спутников «Iridium» весьма велик (достигает 126°). Это приводит, в частности, к тому, что использование на СР антенн размещаемых в одной плоскости оказывается неэффективным. Действительно, всегда найдутся такие направления связи, для которых проекция апертуры антенн на плоскость перпендикулярную этому направлению оказывается сравнительно весьма малой. Поэтому-то и используется 6 антенн, плоскости апертур которых развернуты друг относительно друга, образуя боковую поверхность шестигранной пирамиды, ось симметрии которой совпадает с направлением на подспутниковую точку, причем вершина этой пирамиды обращена к Земле. Ширина диаграммы направленности антенн при последовательном освещении сот изменяется, становясь более узкой по мере увеличения угла направления луча и направлением на подспутниковую точку. Это обеспечивает реализацию равносигнальных диаграмм. Ширины диаграмм направленности лучей СР в L -диапазоне изменяются также в зависимости от текущего положения спутника на орбите. Дело в том, что среднее количество СР видимых с АТ зависит от географического положения АТ: по мере приближения к полюсам Земли это количество возрастает. Чтобы полезно использовать это обстоятельство, по мере приближения СР (в процессе движения по орбите) к полюсам Земли, зоны обслуживания каждого спутника сокращаются за счет сокращения ширины диаграмм направленности всех формируемых им лучей.

С каждого СР системы организуются линии межспутниковой ретрансляции на четыре ближайших СР орбитальной группировки: на СР движущиеся в той же плоскости впереди и

сзади данного, а также на СР движущиеся в плоскостях, лежащих правее и левее данной. Предусматривается возможность многократной ретрансляции сигналов по этим линиям, так что сообщение, поступившее на любой СР, может быть доставлено на любой другой СР орбитальной группировки и передано с него на Землю. Соответствующая маршрутизация осуществляется системой управления, с помощью команд, управляющих коммутацией каналов на СР.

В системе на ЛПВв и ЛПВн, кроме разделения по направлению прихода сигналов, используется еще и частотное и временное разделение. Работа этих линий организуется с циклом, длительность которого составляет 90 мс. Часть этого цикла, длительностью порядка 22 мс, используется для организации симплексных каналов управления системой; остальную его часть занимают восемь слот передачи каналов трафика, вместе с защитными интервалами, разделяющими эти слоты. Каждый слот занимает 8,28 мс. Четыре слота предназначены для работы СР на прием и четыре — на передачу. За время слота борт принимает (передает) один информационный пакет. Соответственно, АТ передают и принимают информацию периодически следующими пакетами.

Что касается использования рабочих частот L -диапазона, то вся выделенная полоса частот шириной 10,5 МГц разделена на 252 стандартные частотные полосы (СЧП), каждая из которых имеет ширину $10\ 500/252 = 125/3 = 41,66$ кГц. При этом, 12 таких СЧП, занимающих верхние 500 кГц выделенного диапазона частот, используются только для передачи симплексных каналов управления. На каждой СЧП передается радиосигнал со скоростью передачи бит $D_{\text{СЧП}} = 50$ кбит/с. При этом используется манипуляция методом ФМ4 со сглаживанием фронтов посылок. Передача ведется с указанным выше временным циклом, так что один СЧП обеспечивает трансляцию четырех каналов по ЛПВв и четырех по ЛПВн. В интересах каждого канала за один временной цикл передается 414 бит ($50\ 000 \times 0,00828$), так что канальная скорость передачи бит составляет 4,6 кбит/с. С учетом избыточности, связанной с кодированием, это и обеспечивает передачу информации со скоростью 2,4 кбит/с. В системе предусматривается и удвоенная скорость передачи информации, для чего соответствующему АТ выделяются два канала связи.

Приведенные данные показывают, что каналы размещены по частоте достаточно плотно, так что внеполосные излучения данного канала влияют на работу соседних. Чтобы сохранить допустимый уровень этих влияний, в системе используется регулировка мощности транслируемых сигналов, поддерживающая этот параметр на минимально необходимом уровне. Далее, для того, чтобы повысить надежность каналов управления, некоторые из них передаются с борта с повышенной мощностью. Так, симплексный канал передачи циркулярных сообщений с борта имеет парциальный ЭИИМ на 15 дБ выше, чем канал трафика, а канал циркулярного вызова — на 6 дБ выше. Соответственно их внеполосные излучения поражают соседние с ними (по частоте) СЧП. Поэтому из 12 СЧП, выделенных для передачи симплексных каналов, используются только 5; остальные являются защитными интервалами (по частоте).

Приведем основные характеристики земных станций системы.

ШС использует три-четыре зеркальные приемо-передающие антенны. Обеспечивается моноимпульсное слежение антеннами за СР. Предусматривается непосредственное взаимодействие ШС с несколькими СР системы, находящимися в данное время в зоне его видимости. Скорость передачи информации на каждой несущей в каждом направлении фидерной линии составляет 3,125 Мбит/с. Мощность передатчика ШС для каждого направления составляет 20 Вт.

АТ представляет собой сочлененные терминалы наземной и спутниковой сотовой связи, благодаря чему может работать в наземной сотовой сети, там где она функционирует. При

отсутствии наземной сети терминал переходит на работу в системе «Iridium». Масса АТ в ручном варианте составляет порядка 0,7 кг, а в возимом — 2,5 кг.

В системе «Iridium» реализована весьма сложная и эффективная система синхронизации. Так, при взаимодействии между СР и АТ обеспечивается синхронный приход пакетов от всех АТ на вход СР, причем тактовая частота манипуляции синхронна с частотой следования пакетов. Далее, при этом взаимодействии, АТ в значительной мере компенсирует нестабильность несущей частоты своего сигнала, принимаемого СР. Если учесть, что для рассматриваемых орбит относительный сдвиг частоты за счет эффекта Доплера достигает величин порядка $(2-3)10^{-5}$, т.е. в абсолютном исчислении порядка 50 кГц, то станет ясной сложность задачи такой синхронизации. При создании системы возникали и многие другие принципиальные проблемы, такие как наведение и слежение антенн межспутниковой связи за соседними спутниками с учетом их взаимного перемещения в процессе движения по орбите, коммутация сигналов отдельных каналов на СР из луча в луч и в линии межспутниковой связи, обеспечение возможности временного разделения между приемом и передачей не только для СР, но и одновременно для АТ с учетом значительного разброса расстояний от различных обслуживаемых АС до СР и др. Разработанные пути решения этих задач являются большим достижением техники связи в целом.

11.3.5. Система «Thuraya»

Региональная система мобильной спутниковой связи «Thuraya» была разработана группой компаний Hughes по заказу ОАР и введена в эксплуатацию в марте 2001 г. Первый спутник-ретранслятор системы был выведен на орбиту близкую к геостационарной (наклонение орбиты 6 градусов) в орбитальную позицию 44° восточной долготы в мае 2000 г. Принципиально зона обслуживания спутника системы «Thuraya» может меняться в полете в зависимости от потребности в ее услугах. В настоящее время зона обслуживания системы охватывает территорию Европы, Центральной Азии и Северной Африки. Всего система оказывает услуги связи в 99 странах, в том числе и в России, на территории от ее западных границ и до Новосибирска на востоке.

Система обеспечивает телефонную связь, обмен телефаксными сообщениями, передачу данных и SMS-сообщений. Обеспечивается как связь подвижных абонентов с использованием АТ системы «Thuraya» с абонентами других систем через ШС и сети общего пользования, так и связь подвижных абонентов между собой «АТ–АТ» с одной ретрансляцией через космический аппарат (КА).

Пропускная способность одного КА «Thuraya» составляет 13 750 дуплексных телефонных каналов. Полностью развернутая система «Thuraya» сможет обслуживать не менее 1750 тыс. абонентов.

Рассмотрим коротко принципы построения системы «Thuraya».

В состав системы «Thuraya» входят один или несколько КА на ГСО, одна или несколько шлюзовых станций, обеспечивающих сопряжение спутниковых каналов системы с каналами связи сетей общего пользования, а также управление коммутацией каналов в ретрансляторе КА и АТ, выполненные в виде телефонной трубки, внешний вид, масса и габаритные размеры которой очень близки к аналогичным параметрам трубок сотовой связи.

Космические аппараты системы «Thuraya» имеют массу на орбите порядка 3200 кг. Солнечные батареи обеспечивают электропитание основных и служебных систем КА мощностью 13,0 кВт. Полезная нагрузка КА (непосредственно ретранслятор) обеспечивает работу в двух диапазонах частот:

- в абонентской линии используется *L*-диапазон: прием (линия АТ–КА) в полосе частот 1626,5–1660,5 МГц, передача (линия КА–АТ) в полосе частот 1525–1559 МГц;
- в фидерной линии используется *C*-диапазон: прием (линия Шлюз–КА) в полосе частот 6425–6725 МГц, линия КА–Шлюз) в полосе частот 3400–3625 МГц.

Для обеспечения почти полуглобальной зоны обслуживания в абонентской линии и возможности использования абонентских терминалов типа трубок (со свойственными трубкам низкими значениями энергетических параметров) в КА «Thuraya» реализована приемо-передающая многолучевая антенна, формирующая 250–300 лучей. Ширина диаграммы направленности каждого луча порядка 1° . Коэффициент усиления 45 дБ. Диаграммы направленности соседних лучей пересекаются по уровню порядка минус (1–2) дБ относительно своего максимума.

Развертываемый уже на орбите рефлектор этой антенны представляет собой несимметричную вырезку из параболоида вращения с размерами малой оси 12,2 м и большой оси 16 м. Диаметр апертуры такого рефлектора составляет 12,2 м. Облучающая система многолучевой антенны выполнена в виде АФАР с числом излучателей порядка 100. Каждый излучатель АФАР участвует в формировании нескольких лучей и, следовательно, каждый луч формируется несколькими излучателями. Принятые разработчиками технические решения по построению облучающей системы позволяют в полете менять зону обслуживания в абонентской линии, а также перераспределять частотный ресурс и излучаемую в каждый луч мощность пропорционально его реальной нагрузке, при этом обеспечивается возможность излучать в сильно загруженный луч до 20% общей излучаемой мощности. Принципы перераспределения частотного ресурса системы и излучаемой мощности между лучами будут кратко рассмотрены ниже.

Методы многостанционного доступа, реализованные в системе «Thuraya» и используемые системой сигналы, определяются стандартом GMR-1 Европейского института стандартизации в области связи (ETSI). Этот стандарт представляет собой спутниковую модификацию стандарта наземной сотовой связи GSM. Так, в соответствии с данным стандартом в абонентской линии системы «Thuraya» используется метод многостанционного доступа сочетающий частотное и временное разделение сигналов. Для передачи используется множество несущих частот, следующих с шагом 31,25 кГц. На каждой такой несущей предусматривается передача с временным делением 24 последовательностей элементарных пакетов (временных слотов), каждый длительностью $5/3$ мс. Период повторения пакетов в каждой последовательности (кадр) составляет соответственно $24 \cdot 5/3 = 40$ мс.

Для формирования канала связи в зависимости от требуемой скорости передачи информации по нему элементарные пакеты могут объединяться. Так, для организации телефонного канала используются пакеты, включающие три идущих подряд элементарных пакета, так что одна несущая при полном ее использовании обеспечивает передачу в одном направлении восьми телефонных каналов. При передаче факсимильной информации и данных используется до девятикратного объединения пакетов. Однократные пакеты могут использоваться в каналах управления системой.

Стандарт GMR-1 предусматривает для пакетов различного назначения использование различных кодов и различных методов модуляции. Это блоковые коды с относительно малой избыточностью и сверточные коды со скоростями $1/2$, $1/3$, $1/4$. Методы модуляции: квадратная фазовая модуляция (QPSK), бинарная фазовая модуляция (BPSK), дифференциальная бинарная фазовая модуляция. Скорость выходного потока модулированных символов на каждой несущей 23,4 ксимв/с. Согласование различных скоростей источников информации (2,4–9,6) кбит/с с постоянной скоростью выходного битового потока обеспечивается использованием кодов с различной избыточностью и пакетов различной

кратности. Применяемые методы демодуляции и декодирования обеспечивают нормальную передачу телефонных сообщений в канале с БГШ при отношении энергии бита сообщения к односторонней спектральной плотности шума порядка $h_6 = 4,5$ дБ.

В соответствии с GMR-1 пять следующих подряд по частоте несущих объединяются в стандартный частотный поддиапазон (СЧПД) шириной 156,25 кГц. Таким образом, при рассматриваемом варианте МСД в СЧПД на длительности одного кадра формируется частотно-временная матрица с $5 \cdot 24 = 120$ позициями.

GMR-1 предусматривает на первом этапе коммутацию каналов; коммутация пакетов планируется в перспективе для более эффективной организации каналов передачи данных и мультимедиа.

Как уже отмечалось в системе «Thuraya», для повышения эффективности использования энергетических и частотных ресурсов обеспечивается гибкое перераспределение СЧПД и ВЧ-мощности между лучами. Перераспределение СЧПД обеспечивается бортовым сигнальным процессором.

Перераспределение излучаемой мощности организуется за счет реализации группового принципа использования УМ в полезной нагрузке КА. Суть этого принципа сводится к объединению нескольких (m) УМ в группу, обслуживающую кластер из m облучателей передающей антенны. УМ группы объединены по входу и выходу соответствующими матрицами, включающими аттенюаторы, фазовращатели, распределители и сумматоры. Сформированная таким образом группа УМ имеет m ВЧ-входов и m ВЧ-выходов. Параметры входной и выходной матриц группы подобраны так, что:

- все УМ группы участвуют в усилении всех входных сигналов этой группы;
- сигнал на j -м входе создает отклик только на j -м выходе, причем до наступления общей перегрузки группа является линейной по каждому входу;
- в выходной матрице осуществляется только когерентное сложение сигналов, так что потери энергии в этой матрице весьма невелики.

В результате реализации указанного принципа использования УМ излучатели, участвующие в формировании лучей с большим трафиком, отбирают у группы УМ большую мощность, чем те, которые связаны с менее загруженными лучами. Таким образом, осуществляется свободное перераспределение ВЧ-мощности между лучами в соответствии с текущим трафиком.

Высокая частотная эффективность достигается за счет реализации семичастотного плана, что обеспечивает при (250–300) лучах примерно (35–40) кратное повторное использование одних и тех же полос частот. Перераспределение частотного ресурса между лучами в системе «Thuraya» осуществляется следующим образом. Методами цифровой фильтрации с помощью мощного бортового процессора в полезной нагрузке реализовано по 600 фильтров на СЧПД как в прямом, так и в обратном направлении. Система формирования частот полезной нагрузки, использующая цифровые вычислительные синтезаторы, построена таким образом, что каждый фильтр может перестраиваться по средней частоте входного и выходного сигналов (независимо) в полосе 34 МГц. Это позволяет использовать такой фильтр для трансляции сигнала СЧПД на любой частоте L -диапазона в любую разрешенную частоту фидерной линии, и обратно. Входные (выходные) сигналы фильтров обратного (прямого) направления, представленные в цифровом виде, могут коммутироваться бортовым цифровым коммутатором в любой луч. Таким образом, и обеспечивается изменение в процессе эксплуатации «своих» полос рабочих частот, их фильтрация, а также распределение и перераспределение выделенного частотного ресурса между лучами. Отметим, что упомянутый бортовой коммутатор обеспечивает, кроме того, переключения на борту, необходимые для организации однокачкового канала при организации связи непосредственно между абонентскими терминалами.

Базовый АТ — портативный абонентский терминал типа «телефонная трубка». Выдвижная антенна терминала выполнена в виде спирали с коэффициентом усиления порядка 3 дБ. ЭИИМ терминала порядка (5–6) дБВт, добротность порядка минус (24–26) дБ/К.

Терминал «Thuraya» является двухрежимным и может работать не только в данной системе, но и в обычной сотовой сети стандарта GSM. Кроме того, в него вмонтирован GPS-приемник, позволяющий определять свои координаты с точностью не менее 100 м. Координаты терминала передаются в служебном сообщении на центральную станцию сопряжения (шлюз). Такое решение позволяет исключать с высокой точностью из зоны обслуживания системой те территории, на которых по тем или иным причинам услуги системы запрещены. Использование GPS-приемников в составе АТ позволяет также поддерживать в терминалах системы единое мировое время, что значительно упрощает процедуры синхронизации их работы в системе.

Терминалы системы «Thuraya» разработаны компанией Hughes Network Systems и производятся двумя производителями оборудования спутниковой связи — Hughes Network Systems и Ascom. По дизайну они незначительно отличаются от обычных сотовых GSM-телефонов. Так, масса терминала составляет 210 г. Для питания АТ используются ионно-литиевые батареи емкостью 650/1200 мА·ч. Максимальное время разговора в режиме спутниковой связи составляет 2,4 ч (батарея на 650 мА·ч) или 4 ч (1200 мА·ч), а в режиме ожидания соответственно 34,1 или 63 ч.

Кроме АТ, в системе будут использоваться также мобильные терминалы. Они включают АТ и набор дополнительных средств, обеспечивающих работу в движении. Максимальная излучаемая мощность передатчика — 2 Вт, т.е. она сопоставима с мощностью мобильных GSM-терминалов. Предполагается также использование терминала типа «таксофона».

Как уже отмечалось, основные функции шлюзовых станций системы «Thuraya»: сопряжение спутниковых каналов системы с каналами связи сетей общего пользования и управление коммутацией каналов в ретрансляторе КА, а также обеспечение биллинга системы и т.п.

В настоящее время в системе функционирует только один «Центральный» шлюз, расположенный в ОАЭ. Шлюз связан с КА «Thuraya» фидерной линией, работающей в С-диапазоне частот.

Оценивая качество предоставляемых системой «Thuraya» услуг, нельзя не отметить, что здесь не удастся обеспечить запаса энергетике в абонентской радиолинии, как в наземных сотовых сетях. Поэтому надежная связь обеспечивается только при наличии линии прямой видимости между терминалом и спутником, т.е. абонент, находящийся в помещении, услуги удовлетворительного качества получить, за редким исключением, не может. Для преодоления неудобств в системе «Thuraya» реализовано следующее техническое решение. Сигнал вызова на спутнике всегда формируется с таким уровнем, которого достаточно, чтобы вызов был принят мобильным абонентом, находящемся, например, в помещении. Приняв вызов, абонент должен создать себе условия для надежной связи.

Список сокращений

АД	Амплитудный детектор	ИКМ	Импульсно-кодовая модуляция
АКФ	Автокорреляционная функция	ИО	Интегральное обслуживание
АМ	Амплитудная манипуляция	КА	Космический аппарат
АСУ-С	Автоматизированная система управления связью	КАМ	Квадратурная амплитудная манипуляция
АТ	Абонентский терминал	КЗЛ	Квазилинейный режим
АТТ	Абонентский терминал телевидения	КК	Квадратурные координаты
АФАР	Активная фазированная антенная решетка	КП	Космическая платформа
АФМ	Амплитудно-фазовая манипуляция	КПСР	Квадратурная псевдослучайная последовательность
АФХ	Амплитудно-фазовая характеристика	ЛБВ	Лампа бегущей волны
АХ	Амплитудная характеристика	ЛМС	Линии межспутниковой связи
АЦП	Аналого-цифровое преобразование	ЛПВв	Линии связи с подвижными объектами вверх
АЧХ	Амплитудно-частотная характеристика	ЛПВн	Линии связи с подвижными объектами вниз
БГШ	Белый гауссовский шум	МДЧР	Множественный доступ с частотным разделением
БЗС	Базовая зонавая станция	МКА	Малый космический аппарат
БИС	Большая интегральная схема	ММС	Манипуляция с минимальным сдвигом
БПФ	Быстрое преобразование Фурье	МС	Местный сигнал
БРС	Базовая радиостанция	МСД	Многостанционный доступ
БС	Базовая станция	МСИ	Межсимвольная интерференция
БТК	Блочные Turboкоды	МСЭ	Международный союз электросвязи
БЧХ	Боуза–Чоудхури–Хоквингема коды	МТП ПРД	Местная тактовая последовательность передатчика
ВИ	Внеполосные излучения	МТП ПРМ	Местная тактовая последовательность приемника
ВКФ	Взаимокорреляционная функция	МШУ	Малошумящий усилитель
ВЧ	Высокочастотная (ый)	МЭ	Мажоритарный элемент
ГАМС	Гармонический ансамбль с минимальным сдвигом	МЭИИ	Множество элементарных исходов испытаний
ГДПСР	Генератор длинной псевдослучайной последовательности	НВ	Неполное вращение
ГКП	Генератор квадратурных переносчиков	НИП	Неравномерная информационная последовательность
ГММС	Гауссовская манипуляция с минимальным сдвигом	НОК	Наименьшее общее кратное
ГС	Групповой сигнал	НСК	Несистематические сверточные коды
ГСО	Геостационарная орбита	НТП	Неравномерная тактовая последовательность
ДПСР	Длинная псевдослучайная последовательность	ОБПФ	Обратное быстрое преобразование Фурье
ДПФ	Дискретное преобразование Фурье	ОБРС	Обучающий радиосигнал
ДРЧ	Двухстороннее расхождение частот	ОКК	Обобщенные каскадные коды
ДСЧП	Дуплексные стандартные частотные полосы	ОКН	Один канал на несущую
ЗО	Зона обслуживания	ОО	Область обслуживания
ЗРК	Запросный радиоканал		
ЗС	Земная станция		

ОРС	Оконечная радиостанция	СЧПД	Стандартный частотный поддиапазон
ОРЧ	Одностороннее расхождение частот	ТК	Турбокоды
ОС	Опорный сигнал	ТРВ	Такты, разрешенные для вставок
ОФМ	Относительная фазовая манипуляция	ТРС	Турбокоды произведения
ПВ	Полное вращение	ТЧ	Тональная частота
ПВЭ	Предельный видеозэквивалент	УГ	Управляемый генератор
ПКБС	Последовательность кадров БС	УМ	Усилитель мощности
ПМК	Последовательность местных кадров	УПС	Узкополосный сигнал
ПН	Полезная нагрузка	УУ	Устройство управления
ППРЧ	Псевдослучайная перестройка рабочей частоты	ФАП	Фазовая автоподстройка
ПРА	Показатель речевой активности	ФАР	Фазированная антенная решетка
ПСП	Псевдослучайная последовательность	ФД	Фазовым дискриминатор
ПУ	Пороговое устройство	ФМ	Фазовая манипуляция
ПФ	Полосовой фильтр	ФМ ШПС	Фазоманипулированный широкополосный сигнал
ПЭС	Поле элементарных событий	ФНЧ	Фильтр нижних частот
РС	Радиостанция	ФПВ	Функционал плотности вероятности
РС	Рида–Соломона коды	ЦАП	Цифро-аналоговый преобразователь
РСК	Решетчатый сигнальный кодер	ЦЗС	Центральная земная станция
РССК	Рекурсивный систематический сверточный код	ЦК	Центр коммутации
САР	Система автоматического регулирования	ЦП	Цифрой преобразователь
СКК	Сигнально-кодовая конструкция	ЦРС	Центральная радиостанция
СО	Сотовое обслуживание	ЦРС	Центральная радиостанция
СП	Спутник-ретранслятор	ЦС	Центральная станция
ССЗ	Система синхронизации по задержке	ЧМ	Частотная манипуляция
ССН	Система синхронизации по несущей	ЧХК	Частотная характеристика канала
ССП	Система синхронизации пакетов	ШПС	Широкополосный, или шумоподобный сигнал
СТК	Сверточные турбокоды	ШРК	Широковещательный радиоканал
СТП	Случайная телеграфная последовательность	ШС	Шлюзовые станции
СТС	Система тактовой синхронизации	ЭВК	Энергетический выигрыш от кодирования
СЦС	Система цикловой синхронизации	ЭИИМ	Эквивалентная изотропно-излучаемая мощность
СЧП	Стандартная частотная полоса	ЭП	Эластичная память
		ЭС	Элементарный сигнал

Литература

Литература к главе 1

- 1.1. Антенно-фидерные устройства и распространение радиоволн / Г.А. Ерохин, О.В. Чернышев, Н.Д. Козырев и др. — М.: Радио и связь, 1996.
- 1.2. *Зыка А.Г. и др. Теория электрической связи. — М.: Радио и связь, 1998.*
- 1.3. *Немировский М.С. Цифровая передача информации. — М.: Связь, 1980.*
- 1.4. *Левин Б.Р. Теоретические основы статистической радиотехники. — М.: Радио и связь, 1989.*

Литература к главе 2

- 2.1. *Шеннон К. Работы по теории информации и кибернетике. — М.: Издательство иностранной литературы, 1963.*
- 2.2. *Яглом А.М., Яглом И.М. Вероятность и информация. — М.: Наука, 1973.*

Литература к главе 3

- 3.1. *Котельников В.А. Теория потенциальной помехоустойчивости. — М.: Госэнергоиздат, 1956.*
- 3.2. *Финк Л.М. Теория передачи дискретных сообщений. — М.: Сов. радио, 1970.*
- 3.3. *Маковеева М.М., Шинаков Ю.С. Системы связи с подвижными объектами. — М.: Радио и связь, 2002.*

Литература к главе 4

- 4.1. *Шеннон К.Е. Математическая теория связи / Работы по теории информации и кибернетике. — М., 1963.*
- 4.2. *Питерсон У., Уелдон Э. Коды, исправляющие ошибки: Пер. с англ. / Под ред. Р.Д. Добрушина и С.И. Самойленко. — М.: Мир, 1976.*
- 4.3. *Берлекэмп Э. Алгебраическая теория кодирования: Пер. с англ. / Под ред. С.Д. Бермана. — М.: Мир, 1971.*
- 4.4. *Прокис Дж. Цифровая связь: Пер. с англ. / Под ред. Д.Д. Кловского. — М.: Радио и связь, 2000.*
- 4.5. *Колесник В.Д., Мирончиков Е.Т. Декодирование циклических кодов. — М.: Связь, 1968.*
- 4.6. *Кларк Дж., Кейн Дж. Кодирование с исправлением ошибок в системах цифровой связи: Пер. с англ. — М.: Радио и связь, 1987.*
- 4.7. *Блейхут Р. Быстрые алгоритмы цифровой обработки сигналов: Пер. с англ. — М.: Мир, 1989.*
- 4.8. *Теория кодирования: Пер. с японск. / Т. Кассама, Н. Токура, Е. Ивадари и др.; Под ред. Б.С. Цыбакова и Гельфанда. — М.: Мир, 1978.*
- 4.9. *Банкет В.Л., Дорофеев В.М. Цифровые методы в спутниковой связи. — М.: Радио и связь, 1988.*
- 4.10. *Варакин Л.Е. Теория систем сигналов. — М.: Сов. радио, 1978.*
- 4.11. *Левин Б.Р. Теоретические основы статистической радиотехники. Кн. первая. 2-е изд., перераб. и доп. — М.: Сов. радио, 1974.*
- 4.12. *Зяблов В.В., Коробков Д.Л., Портной С.Л. Высокоскоростная передача сообщений в реальных каналах. — М.: Радио и связь, 1991.*
- 4.13. *Спилкер Дж. Цифровая спутниковая связь: Пер. с англ. / Под ред. В.В. Маркова. — М.: Связь, 1979.*
- 4.14. *Мессис Дж. Пороговое декодирование. — М.: Мир, 1966.*
- 4.15. *Viterby A.J. Error bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm, IEEE Trans. Info. Theory, Apr. 1967.*
- 4.16. *Фано Р. Передача информации. Статистическая теория связи: Пер. с англ. / Под ред. Р.Л. Добрушина. — М.: Мир, 1965.*
- 4.17. *Золотарев В.В. Реальный энергетический выигрыш кодирования для спутниковых каналов. — В кн.: 4-я Международная конференция «Спутниковая связь-2000». Том 2. — М., 2000.*
- 4.18. *Витерби А.Д., Омара Д.К. Принципы цифровой связи и кодирования: Пер. с англ. / Под ред. Зигангирова. — М.: Радио и связь, 1982.*
- 4.19. *Форни Д. Каскадные коды. — М.: Мир, 1970.*

- 4.20. *Berrou C., Glavieux A., Thitimajshima P.* Near Shannon Limit Error-Correcting Coding and Decoding: Turbo-Codes // Proceeding of ICC'93, Geneva, Switzerland. P. 1064–1070, May 1993.
- 4.21. *Dolinar S., Divsalar D.* Weight Distribution for Turbo Codes Using Random and Nonrandom Permutation / TDA Progress Report 42-122, Jet Propulsion Laboratory, Pasadena, California, August, 15. 1995.
- 4.22. *Bahl L.R., Cocke J., Jelinek F., Raviv J.* Optimal decoding of linear codes for minimizing symbol error rate. IEEE Trans, Inform Theory. V. IT-20. March 1974.
- 4.23. *Скляр Б.* Цифровая связь. Теоретические основы и практическое применение. — 2-е изд., испр.: Пер. с англ. — М.: Издательский дом «Вильямс», 2003.

Литература к главе 5

- 5.1. *Банкет В.Л., Дорофеев В.М.* Цифровые методы в спутниковой связи. — М.: Радио и связь, 1988.
- 5.2. *Зяблов В.В., Коробков Д.Л., Портной С.Л.* Высокоскоростная передача сообщений в реальных каналах. — М.: Радио и связь, 1991.
- 5.3. *Прокис Дж.* Цифровая связь. — М.: Радио и связь, 2000.

Литература к главе 6

- 6.1. *Немировский М.С.* Цифровая передача информации в радиосвязи. — М.: Связь, 1980.
- 6.2. *Бернард Скляр.* Цифровая связь. Теоретические основы и практическое применение. — М.: СПб., Киев: Вильямс, 2003.
- 6.3. *Маковеева М.М., Шинаков Ю.С.* Системы связи с подвижными объектами. — М.: Радио и связь, 2002.

Литература к главе 7

- 7.1. *Петрович Н.Т., Размахнин М.К.* Системы связи с шумоподобными сигналами. — М.: Сов. радио, 1969.
- 7.2. *Варакин Л.Е.* Теория сложных сигналов. — М.: Сов. радио, 1970.
- 7.3. *Варакин Л.Е.* Теория систем сигналов. — М.: Сов. радио, 1978.
- 7.4. *Варакин Л.Е.* Системы связи с шумоподобными сигналами. — М.: Радио и связь, 1985.
- 7.5. *Питерсон У., Уелдон Э.* Коды, исправляющие ошибки: Пер. с англ. / Под ред. Р.Д. Добрушина и С.И. Самойленко. — М.: Мир, 1976.
- 7.6. *Сандерс.* Система связи «Диджилок». — В кн.: Передача цифровой информации: Пер. с англ.; Под ред. С.И. Самойленко. — М.: ИЛ, 1963.
- 7.7. *Gold R.* Maximal Recursive sequences with 3-valued Recursive Cross-Correlation Functions // IEEE Trans. Inf. Th.. 1968. V. IT-14. № 1. P. 154–156.
- 7.8. *Кор, Кручфилд, Мерчиз.* Импульсная УКВ станция, использующая шумоподобные сигналы // Зарубежная радиоэлектроника. — 1966. — Т. 20. — № 4.
- 7.9. *Gilhausen K.S. et al.* Increased Capacity Using CDMA for Mobile Satellite Communication // IEEE Journal on Selected Areas in Communications. 1990. V. 8. № 4. P. 503–513.
- 7.10. *Viterbi A.J.* A Perspective on the Evolution of Multi-Access Satellite Communication // IEEE Journal on Selected Areas in Communications. 1992. V. 10. № 6. P. 980–983.
- 7.11. Помехозащищенность радиосистем со сложными сигналами / Г.И. Тузов, В.А. Сивов, В.И. Прытков и др.; Под ред. Г.И. Тузова. — М.: Радио и связь, 1985.
- 7.12. *Viterbi A.J., Jacobs I.M.* Advances in coding and modulation for noncoherent channels affected by fading, partial band and multiple-access interference // Advances in Communication Systems. 1975. № 4.
- 7.13. *Биленко А.П., Волков Л.Н.* Сравнение помехозащищенности радиолиний с широкополосными сигналами // Радиотехника. — 1986. — № 4.
- 7.14. *Комаров В.Н.* Об исправлении многократных пакетов ошибок // Техника средств связи. — 1980. — Вып. 3.

Литература к главе 8

- 8.1. *Rappaport T.S.* Wireless Communications (Principles and Practice). — N.Y.: Prentice Hall PRT, 1996.
- 8.2. *Lee W.C.Y.* Mobile Cellular Telecommunications Systems. — N.Y.: McGraw Hill Publications, 1989.

- 8.3. *Smith J.I.* A Computer Generated Multipath Fading Simulation for Mobile Radio // IEEE Transactions on Vehicular Technology. V. VT-24. № 3. August 1975. P. 39–40.
- 8.4. Annex C (normative): Propagation conditions. — ETSI: EN 300910 V8.5.1 (GSM 05.05, version 8.5.1 Release 1999) (2000–11).
- 8.5. *Dexiang Meng, Weiling Wu.* A Model of Mobile Satellite Channel // Proceedings of ICCT-2003. P. 1067–1069.
- 8.6. GMR-1 GEO Mobile Radio — family-1. ETSI TS 101 376-5-5, v1.1.1 (2001–03). Part 5: Radio interference physical layer specifications. Sub-part 5: Radio Transmission and Reception.
- 8.7. Signal Processing for Mobile Communications handbook / Edited by M. Ibnkahla. — N.Y.: CRC Press, 2004.
- 8.8. *Li W., Law C.L., Dubey V.K., Ong J.T.* Ka-band land mobile satellite channel model incorporating weather effects // IEEE Comm. Letters, 2001. V. 5. № 5. P. 194–196.
- 8.9. *Fontan F.P., Vazquez-castro M., Cabado C.E., Garsia J.P., Kubista E.* Statistical modeling of the LMS channel // IEEE Trans. Veh. Technol. 2001. V. 50. № 11. P. 1549–1567.
- 8.10. *Burgueno A., Vilar E., Puigcerver M.* The land mobile satellite communication channel — Recording, statistics and channel model // IEEE trans. Veh. Technol. 1991. Vol. 40. № 5. P. 375–384.

Литература к главе 9

- 9.1. *Финк Л.М.* Теория передачи дискретных сообщений. — М.: Сов. радио, 1970.
- 9.2. *Левин Б.П.* Теоретические основы статистической радиотехники. — М.: Сов. радио, 1975. — Кн. 2.
- 9.3. *Bellman R.E.* Dynamic Programming. — Princeton, N.J.: Princeton University Press, 1957.
- 9.4. *Viterbi A.J.* Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm // IEEE Trans. Info. Theory. V. IT-13. Apr. 1967. P. 260–269.
- 9.5. *Chang R.W., Hancock J.C.* On Receiver Structures for Channels Having Memory // IEEE Trans. Info. Theory. V. IT-12. Oct. 1966. P. 463–468.
- 9.6. *Forney G.D.* Maximum Likelihood Sequence Estimation of Digital Sequences in the Presence of Inter-symbol Interference // IEEE Trans. Info. Theory. V. IT-18. May 1972. P. 363–378.
- 9.7. *Hayes J.F.* The Viterbi Algorithm Applied to Digital Data Transmission // IEEE Communications Magazine. May 2002. P. 26–32.
- 9.8. *Rappaport T.S.* Wireless Communications. Principles and Practice. — N.Y.: IEEE Press, 1996.
- 9.9. Signal Processing for Mobile Communications Handbook / Edited by M. Ibnkahla. — London: CRC Press, 2004.
- 9.10. *Sang-Sung Han et al.* New Soft-Output MLSE Equalization Algorithm for GSM Digital Cellular Systems // IEEE. 0-7803-2956-4/95.
- 9.11. *Nee R. van, Prasad R.* OFDM for wireless multimedia communications. — N.Y.: Artech House, 2000.
- 9.12. *Рабинер Л., Гоулд Б.* Теория и применение цифровой обработки сигналов: Пер с англ. / Под ред. Ю.Н. Александрова. — М.: Мир, 1978.

Литература к главе 10

- 10.1. *Немировский М.С.* Цифровая передача информации в радиосвязи. — М.: Связь, 1980.
- 10.2. *Прокис Дж.* Цифровая связь. — М.: Радио и связь, 2000.
- 10.3. *Шахгильдян В.В., Ляховкин А.А.* Системы фазовой автоподстройки частоты. — М.: Связь, 1972.