

В. М. Ближневский, А. И. Ляков,
С. А. Портной, И. В. Шахнович

ШИРОКОПОЛОСНЫЕ БЕСПРОВОДНЫЕ СЕТИ ПЕРЕДАЧИ ИНФОРМАЦИИ



—• Wi-Fi с активной

Оборудование Wi-Fi с активной многолучевой ФАР компании Vivato (США) предназначено для построения сетей Wi-Fi большой площади. Площадь покрытия одной базовой станции Vivato может составлять несколько десятков квадратных километров.

• Области применения

Крупные хот-споты в городах (улицы, площади, пляжи), сети доступа в Интернет (загородные поселки, студенческие городки, сельская связь), сети технологической связи (железнодорожные станции и терминалы, морские порты, аэропорты, таможенные терминалы, промышленные районы), беспроводные LAN в помещениях (офисы, склады, заводские цеха, аэропорты, супермаркеты и т. п.).



• Интеллектуальная коммутация

Встроенный коммутатор имеет два магистральных порта и по одному Wi-Fi порту на каждое из возможных направлений луча ФАР (6 или 13). Передача и прием данных осуществляются по пакетно по патентованной технологии PacketSteering. Например, когда приходит пакет из магистральной, коммутатор передает его на соответствующий Wi-Fi порт, ФАР формирует "луч" в требуемом направлении на время передачи пакета и пакет передается абоненту.

многолучевой ФАР. —



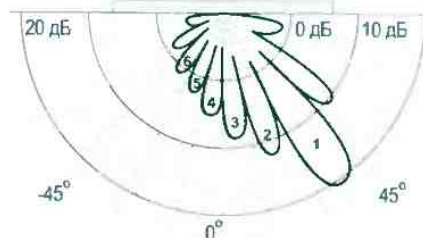
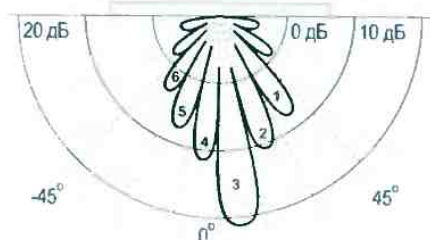
VP1200 - базовая станция 802.11b
для помещений



VP2210 - базовая станция 802.11b,g
в термобоксе (до -40°C)

● "Умная" антенна

В качестве антенны используется активная многолучевая фазированная антенная решетка (ФАР). Антенна имеет несколько возможных направлений "луча" (основного лепестка диаграммы направленности) - 6 или 13. Формирование "луча" в любом направлении осуществляется за несколько микросекунд. Образно говоря, антенна сканирует широкий сектор покрытия узким направленным лучом, обеспечивая большую дальность связи.



Диаграммы направленности. Вид сверху.
Время переключения "луча" - несколько микросекунд

● Характеристики

	VP1200, VP1210	VP2200, VP2210
Стандарт	802.11b	802.11b,g
Частотный диапазон	2,4-2,483 ГГц	2,4-2,483 ГГц
Пропускная способность	5-10 Мбит/с	5-20 Мбит/с
Число активных абонентов	до 100	до 50
Антенна	активная ФАР 24 dBi	активная ФАР 21 dBi
сектор покрытия	100° x 12°	90° x 12°
число направлений луча	13	6
прием (одновременно)	все направления	все направления
передача (одновременно)	1 или 2 направления	1 направление
Дальность связи (на стандартный адаптер с антенной 2,2 dBi в прямой видимости)		
1 Мбит/с	более 7 км	более 7 км
11 Мбит/с	более 4 км	более 4 км
54 Мбит/с	-	более 600 м
Безопасность	WEP + 802.1x	802.11i, WPA, WEP
Магистраль	Fast и Gigabit Ethernet	Fast Ethernet

РОССИЙСКАЯ АКАДЕМИЯ НАУК
ИНСТИТУТ ПРОБЛЕМ ПЕРЕДАЧИ ИНФОРМАЦИИ

**В. М. Вишневский, А. И. Ляхов,
С. Л. Портной, И. В. Шахнович**

ШИРОКОПОЛОСНЫЕ БЕСПРОВОДНЫЕ СЕТИ ПЕРЕДАЧИ ИНФОРМАЦИИ

ТЕХНОСФЕРА
МОСКВА
2005



*Издание осуществлено при финансовой поддержке
Российского Фонда Фундаментальных Исследований
по проекту № 05-07-95003*

В.М. Вишневский, А.И. Ляхов, С.Л. Портной, И.В. Шахнович
Широкополосные беспроводные сети передачи информации
Москва:
Техносфера, 2005 — 592 с.

В монографии описаны принципы построения, логическая и физическая структуры беспроводных сетей передачи информации, включая персональные сети (IEEE 802.15, в том числе стандарт Bluetooth), локальные беспроводные сети (стандарты IEEE 802.11 и DECT), региональные сети стандарта IEEE 802.16, сотовые телефонные и транкинговые сети, системы цифрового теле- и радиовещания, спутниковые системы и оптические атмосферные каналы связи. Изложены теоретические основы передачи информации (теоремы Шеннона, Котельникова и Найквиста), методы кодирования и модуляции в беспроводных сетях. Описаны новые методы оценки производительности и проектирования локальных и региональных беспроводных сетей под управлением протоколов IEEE 802.11(16). Рассмотрены принципы построения и реализация антенных систем для широкополосной связи. Описаны основы схемотехники и принципы функционирования smart-антенн и технологии MIMO, положенной в основу стандартов IEEE 802.11n и IEEE 802.16-2004. Приведены примеры реализации региональных сетей; описана новая технология построения широкополосных беспроводных сетей с использованием высотных платформ.

© 2005, В.М. Вишневский, А.И. Ляхов,
С.Л. Портной, И.В. Шахнович
© 2005, ЗАО «РИЦ «Техносфера»
оригинал-макет, оформление.

ISBN 5-94836-049-0

ОГЛАВЛЕНИЕ

Предисловие	10
Введение	12
1 Беспроводные сети передачи информации.	
Основные понятия	17
1.1 Исторический очерк развития сетевых технологий	17
1.2 Классификация и технологии беспроводных сетей	24
1.3 Стандартизация в области телекоммуникаций	27
1.4 Модель взаимодействия открытых систем	30
1.5 Методы доступа к среде в беспроводных сетях	32
2 Технологии и архитектура беспроводных сетей	39
2.1 Персональные беспроводные сети (технологии Bluetooth, Home RF, IEEE 802.15.3(4))	39
2.1.1 Стандарты Bluetooth и HomeRF	39
2.1.2 Архитектура и логическая структура сети Bluetooth	40
2.1.3 Технические средства сети Bluetooth	45
2.1.4 Высокоскоростные персональные сети стандарта IEEE 802.15.3(3a)	48
2.1.5 Низкоскоростные сети стандарта IEEE 802.15.4 (ZigBee)	53
2.1.6 Технология сверхширокополосной связи	56
2.2 Беспроводные локальные сети (стандарты DECT и IEEE 802.11)	64
2.2.1 Локальные сети под управлением IEEE 802.11	64
2.2.2 Стандарт DECT	70
2.3 Беспроводные сети регионального масштаба	80
2.3.1 Региональные сети широкополосного доступа под управле- нием IEEE 802.16	82
2.3.2 Мобильные сотовые технологии	83
2.3.3 Технологии транковой радиосвязи	98
2.3.4 Широковещательные сети — цифровое телевидение	110
2.3.5 Широковещательные сети — цифровое радио	123
2.4 Спутниковые сети	138

2.4.1	Исторический очерк развития спутниковой связи	138
2.4.2	Виды орбитальных группировок. Геостационарные орбиты	140
2.4.3	Эллиптические, средневысотные и низкие орбиты	141
2.4.4	Архитектура и основные принципы работы спутниковых систем связи	143
2.4.5	Методы множественного доступа в ССС	145
2.5	Оптические атмосферные линии связи и сети	147
3	Математические основы передачи информации	153
3.1	Структура и функциональная схема систем передачи дискретной информации. Описание элементов систем (кодеки, модемы, каналы)	153
3.2	Модели и типы каналов широкополосного доступа	155
3.3	Система передачи дискретной информации как система широкополосного доступа к информационным ресурсам. Практические примеры систем	156
3.4	Аналоговая и цифровая передача информации	157
3.4.1	Информация как физическая величина	157
3.4.2	Энтропия Л. Больцмана	158
3.4.3	Определение информации для дискретных систем связи	160
3.4.4	Безусловная энтропия	161
3.4.5	Условная энтропия	163
3.4.6	Взаимная информация	166
3.4.7	Свойства энтропии и информации	168
3.4.8	Пропускная способность дискретного канала	169
3.4.9	Скорость передачи	170
3.4.10	Аналоговые источники	170
3.4.11	Максимизация энтропии	171
3.5	Теорема отсчетов и ее использование для выбора режимов функционирования канала связи	172
3.5.1	Теорема отсчетов	172
3.5.2	Теорема отсчетов для сигналов с ограниченной полосой	174
3.5.3	Теорема 1 Котельникова. Сигнал $s(t)$ задан на бесконечной оси времени, $-\infty < t < +\infty$	175
3.5.4	Физический смысл теоремы 1 Котельникова	175
3.5.5	Необходимое уточнение теоремы 1 Котельникова	178
3.5.6	Можно ли переносить теорему на случайные сигналы?	179
3.5.7	Что получится, если в теореме Котельникова отсчеты брать чаще или реже, чем с периодом $\tau = 1/F$?	182
3.5.8	Ряд Котельникова при конечном числе отсчетов k	183
3.5.9	Теорема отсчетов в частотной области (теорема 2)	185
3.5.10	Обобщение теоремы Котельникова на случай произвольной полосы частот ширины F	186
3.5.11	Теорема отсчетов для АМ и ЧМ сигналов	188
3.5.12	Неравномерные отсчеты	189
3.5.13	Теорема для сгруппированных отсчетов	190

3.5.14	Теорема для неравномерных достаточно регулярных отсчетов	191
3.6	Теоремы Найквиста и Шеннона	192
3.6.1	Формула Найквиста	192
3.6.2	Обобщения формулы Найквиста	197
3.6.3	Формула Шеннона–Таллера–Хартли	198
3.6.4	Теоремы Шеннона для сигналов с ограниченным спектром	199
3.6.5	Теоремы Шеннона для дискретных сообщений	200
3.6.6	Теорема Шеннона в терминах полезной информации	200
4	Коды и их применение в системах передачи информации	203
4.1	Коды, устраняющие избыточность	203
4.1.1	Введение в теорию кодирования	203
4.1.2	Теорема Шеннона для дискретного источника	205
4.1.3	Применение кодов, устраняющих избыточность	206
4.2	Общее понятие о шифровании информации	213
4.3	Корректирующие коды	215
4.3.1	Блок-схема системы связи и примеры простейших кодов	215
4.3.2	Теорема Шеннона для канала с шумами	218
4.3.3	Введение в теорию групп, колец и полей	219
4.3.4	Введение в пространства Хемминга	226
4.3.5	Линейные коды	228
4.3.6	Циклические коды	233
4.3.7	Наиболее известные классы блочных кодов	235
4.3.8	Итеративные и каскадные коды	238
4.3.9	Мягкое декодирование, энергетический выигрыш кодирования — основные определения	239
4.3.10	Низкоплотностные коды	241
4.3.11	Сверточные коды	246
4.3.12	Турбокоды	251
4.3.13	Обобщенные каскадные коды	255
5	Системы модуляции и сигнально-кодовые конструкции	258
5.1	Модуляция как перенос сигнала по спектру	258
5.2	Дискретная модуляция	259
5.3	Сигнально-кодовые конструкции (СКК) в гауссовом канале	260
5.4	Описание блочных СКК в гауссовском канале	263
5.5	Описание сверточных СКК в гауссовом канале	265
5.6	Модель канала с межсимвольной интерференцией (МСИ)	265
5.7	Преобразование канала с МСИ в параллельные каналы без памяти	269
5.8	Пропускная способность канала с МСИ	273
5.9	Построение СКК для канала с МСИ и переменными параметрами (OFDM)	274

6 Беспроводные локальные сети стандарта IEEE 802.11 и оценка их производительности	277
6.1 Основные принципы IEEE 802.11	277
6.2 MAC-уровень стандарта IEEE 802.11	280
6.3 Физический уровень стандарта IEEE 802.11b	282
6.4 Аппаратная реализация	286
6.5 Стандарты IEEE 802.11a и 802.11g	289
6.6 Схема распределенного управления в локальных сетях	295
6.7 Обобщение существующих методов оценки производительности БЛС	298
6.8 Эффект захвата	301
6.8.1 Описание эффекта захвата	301
6.8.2 Оценка виртуального времени передачи	302
6.8.3 Оценка вероятностей передачи и мгновенного повторения попытки передачи для обычных станций	306
6.9 Численные результаты оценки производительности в условиях идеального канала	311
6.10 Производительность локальных сетей при наличии помех	314
6.10.1 Оценка пропускной способности	315
6.10.2 Вероятности начала передачи и отказа	318
6.11 Результаты оценки производительности сети с помехами	324
6.11.1 Оптимизация порога RTS и порога фрагментации	324
6.11.2 Сравнительный анализ правил отсрочки	327
6.12 Производительность сети при коррелированных сбоях	329
6.12.1 Пропускная способность при коррелированных сбоях	329
6.12.2 Успешный слот	331
6.12.3 Алгоритм вычисления вероятностей	333
6.12.4 Численные результаты	336
6.13 Производительность локальных сетей при нормальной нагрузке	338
6.13.1 Модель Маркова	339
6.13.2 Оценка показателей производительности	345
6.14 Моделирование широкополосной беспроводной сети	348
6.14.1 Аналитическая модель	348
6.14.2 Оценка производительности сети	352
6.14.3 Имитационная модель беспроводной сети	353
6.15 Оценка пропускной способности беспроводной сети типа «hot spot» и ее оптимизация	357
6.15.1 Сети типа «hot spot»	357
6.15.2 Математическая модель сети	359
6.15.3 Оптимизация пропускной способности	363
6.15.4 Численные оценки	365
7 Оценка производительности городских и региональных беспроводных сетей	370
7.1 Работа протокола IEEE 802.11 в условиях городских и региональных сетей	370

7.2	Моделирование радиосоты	373
7.2.1	Имитационное моделирование радиосоты	373
7.2.2	Аналитический метод оценки пропускной способности	375
7.2.3	Метод оценки пропускной способности при технологии FHSS	380
7.3	Численные результаты исследования городской радиосоты	382
7.4	Оценка максимальной производительности беспроводного доступа в Интернет	387
7.4.1	Оценка пропускной способности	388
7.4.2	Анализ насыщенной сети	391
7.4.3	Численные результаты	395
7.5	Применение централизованного управления	397
7.6	Динамический опрос в беспроводных сетях с централизованным управлением	401
7.6.1	Стратегии опроса	401
7.6.2	Аналитическая модель динамической стратегии отложенно- го опроса	402
7.6.3	Оценка показателей производительности	408
7.6.4	Численные результаты	410
7.7	Экспериментальные исследования и оптимизация городской бес- проводной сети	415
7.7.1	Способы получения и усреднения статистических данных	417
7.7.2	Анализ полученных данных	419
7.7.3	Постановка задачи оптимизации	422
7.7.4	Метод оптимизации	425
7.7.5	Результаты расчетов	427
7.8	Математические методы топологического проектирования широко- полосных беспроводных сетей	432
7.8.1	Постановка задачи оптимального размещения базовых стан- ций и подключения к ним локальных компьютерных сетей абонентов	432
7.8.2	Алгоритм решения задачи псевдобулевой оптимизации	436
8	Стандарт широкополосного доступа IEEE 802.16-2004	444
8.1	MAC-уровень стандарта IEEE 802.16	446
8.1.1	Структура MAC-уровня	446
8.1.2	Соединения и сервисные потоки	447
8.1.3	Пакеты MAC-уровня	448
8.1.4	Общая структура кадров IEEE 802.16	450
8.1.5	Принцип предоставления канальных ресурсов	451
8.1.6	Механизмы подтверждения приема и быстрой обратной связи	452
8.2	Физический уровень стандарта IEEE 802.16	453
8.2.1	Режим WirelessMAN-SC	454
8.2.2	Режим WirelessMAN-OFDM	458
8.2.3	Mesh-сеть	465
8.2.4	Режим OFDMA	469

8.3	Поддержка адаптивных антенных систем	475
8.3.1	Работа с направленными ААС	476
8.3.2	Пространственно-временное кодирование	478
8.4	Аппаратная поддержка стандарта IEEE 802.16	480
8.4.1	Интегральная элементная база	480
8.4.2	Особенности реализации аппаратуры стандарта IEEE 802.16	482
8.5	Будущее широкополосного беспроводного доступа по стандартам IEEE 802.16	483
9	Архитектура и технические средства беспроводных региональных сетей	486
9.1	Архитектура беспроводных региональных сетей с фиксированной топологией	486
9.2	Логическая и аппаратная структура радиомаршрутизатора «РА-ПИРА»	489
9.2.1	Аппаратная платформа	490
9.2.2	Проблема частотного ресурса	491
9.2.3	Логическая схема абонентского тракта	492
9.2.4	Обеспечение информационной безопасности	494
9.2.5	Температурная стабилизация в РЭС «РАПИРА»	496
9.2.6	Использование технологии PoE (Power over Ethernet)	497
9.3	Фидерные линии и элементы СВЧ-тракта	498
9.3.1	Основные понятия	498
9.3.2	Двухпроводные и коаксиальные линии	500
9.3.3	Прямоугольные и круглые волноводы	501
9.3.4	Направленные ответвители и волноводные мосты	503
9.3.5	Ферритовые фазовращатели и циркуляторы	505
9.4	Основные понятия теории и техники антенн. Классификация антенных систем	507
9.5	Антенные системы евклидовой геометрии	514
9.5.1	Элементарные антенны (проволочные вибратор и рамка)	514
9.5.2	«Ненаправленные» антенны	516
9.5.3	Направленные антенны	517
9.6	Фрактальные антенны	529
9.6.1	Излучатель на основе фрактала Коха	529
9.6.2	Диполь на основе фрактала Минковского	534
9.6.3	Древовидные диполи	534
9.6.4	Рамочные антенны на основе фрактальных петель Коха и Минковского. Крест Минковского	536
9.6.5	Монополь, ковер и куб Серпинского. Проволочные трехмерные фракталы	539
9.7	SMART-антенны. Цифровые антенные решетки (ЦАР)	542
9.7.1	Понятие ЦАР. Преимущества ЦАР по сравнению с ФАР	542
9.7.2	Методы пространственной селекции сигналов и подавления помех в ЦАР	553

9.8	ММО-системы на базе ЦАР	556
9.9	Особенности построения суперлинейных усилителей	564
9.10	Построение беспроводных сетей передачи информации с использо- ванием привязных аэростатов	569
9.10.1	Классификация технологий высотных платформ	570
9.10.2	Преимущества беспроводных сетей на основе привязных аэростатов	572
9.10.3	Архитектура и проблемы реализации	573
9.10.4	Проект БАРС	575
Литература		579

Предисловие

В последнее десятилетие XX века беспроводные цифровые коммуникации вступили в фазу бурного развития, которая продолжается и в настоящее время. Толчком к этому послужило, с одной стороны, начавшееся интенсивное развитие глобальной сети Интернет, с другой — внедрение новых, прогрессивных методов кодирования, модуляции и передачи информации. В настоящее время очевидно, что беспроводные широкополосные сети практически находятся вне конкуренции по оперативности развертывания, мобильности, цене и широте возможных приложений, во многих случаях представляя собой единственное экономически оправданное решение.

Для стран, в которых большая территория сочетается с невысокой плотностью населения, широкополосные беспроводные решения имеют особое значение, так как позволяют экономично и оперативно создавать телекоммуникационную инфраструктуру на обширных территориях. Особенно важно это для информатизации удаленных и сельских регионов Российской Федерации и решения одной из важнейших проблем информационной безопасности России — проблемы «информационного неравенства» российских регионов.

В связи с этим весьма актуальной является дальнейшая разработка фундаментальной теории в области передачи информации, методов проектирования, алгоритмов и отечественных программно-технических средств комплектации широкополосных беспроводных сетей (высокоскоростных радиомодемов и управляемых антенных устройств). Все эти вопросы нашли полное отражение в настоящей книге, что определяет своевременность и актуальность ее появления.

В книге изложены теоретические основы передачи информации (теоремы Шеннона, Котельникова и Найквиста), методы кодирования и модуляции в беспроводных сетях. Описаны новые методы и модели оценки производительности и проектирования локальных и региональных беспроводных сетей под управлением протоколов IEEE 802.11(16).

Дано описание принципов построения, логической и физической структуры беспроводных сетей передачи информации, включая персональные сети (стандарты Bluetooth, IEEE 802.15.1 (3, 3а, 4)), локальные беспроводные сети (стандарты IEEE 802.11 и DECT), региональные сети стандарта широкополосного вещания IEEE 802.16, сотовые телефонные и транкинговые сети, системы цифрового теле- и радиовещания, спутниковые системы и оптические атмосферные каналы связи.

Рассмотрены принципы построения и реализация антенных систем широкополосных систем связи. Описаны основы разработки схемотехники и принципы

функционирования smart-антенн и базирующейся на них технологии ММО, положенной в основу широкополосных стандартов IEEE 802.11n и IEEE 802.16-2004. Приведены примеры реализации региональных сетей; описана новая технология построения широкополосных беспроводных сетей с использованием высотных платформ.

Авторы книги являются известными специалистами в области проектирования и реализации сетей передачи информации. В основу книги положены полученные ими оригинальные теоретические результаты и опыт создания локальных и крупномасштабных широкополосных беспроводных сетей.

Надеемся, что настоящая монография будет полезна широкому кругу читателей: студентам и аспирантам, научным работникам и инженерам-проектировщикам, занимающимся фундаментальными и прикладными исследованиями в области телекоммуникационных сетей.

Академики РАН Ю.В. Гуляев, Н.А. Кузнецов

Введение

В последние годы беспроводные сети передачи информации становятся одним из основных направлений развития телекоммуникационной индустрии. Однако отметим, что беспроводная передача информации существует столько же, сколько существует человеческая цивилизация. Гонцы, стрелы, сигнальные костры, телеграф, искровые передатчики, спутниковые системы связи — все это звенья одной цепи. Изменялись технологии, но суть сетей передачи оставалась неизменной — организовать взаимодействие нескольких различных элементов так, чтобы информация в заданное время поступала из одной точки в другую. Исторически цифровая (дискретная) передача на шаг опережала аналоговую (письмо написать проще, чем сохранить на каком-либо носителе речь). В конце XIX века окончательно оформились две ветви единого телекоммуникационного дерева — передача голоса (телефония) и данных (телеграф). Примерно в то же время началось разделение на проводные и беспроводные технологии передачи. Проводная связь в ту эпоху оказалась проще, надежнее, защищеннее. Начался век проводных телекоммуникаций. Тысячи километров кабелей опутали Землю, как паутина. Человечество потребляло все больше информации и все больше увязало в путях медной проволоки и кварцевого оптического волокна.

К концу XX века в технологии связи возникла новая волна — цифровая обработка. Вскоре практически любую информацию перед трансляцией, будь то речь или телевизионная картинка, стали преобразовывать в поток нулей и единиц. Настала эпоха цифровой связи. Благодаря цифровой обработке начали все теснее переплетаться развивавшиеся параллельно технологии телефонии и передачи данных, чтобы с появлением пакетных сетей слиться практически воедино. Появился даже термин «мультимедиа», означающий объединение самых различных информационных технологий (голос, аудио/видео, данные) в единой технологической среде обработки и передачи. Взрывоподобное развитие Интернета лишь подтвердило тот факт, что цифровые сети передачи для современной цивилизации стали столь же необходимы, как автостреды, трубопроводы и линии электропередачи.

Локальные и региональные сети проникли во все сферы человеческой деятельности, включая экономику, науку, культуру, образование, промышленность и т.д. Технологию Ethernet (10 Мбит/с) сменили Fast Ethernet/Gigabit Ethernet (100/1000 Мбит/с), в глобальных сетях свершился переход от негорючей, но сверхнадежной технологии X.25 к методу Frame Relay, применению стека протоколов TCP/IP, к технологиям ATM и GigaEthernet. Без них невозможны столь привычные сегодня электронная почта, факсимильная и телефонная связь, до-

ступ к удаленным базам данных в реальном масштабе времени, службы новостей, дистанционное обучение, телемедицина, телеконференции, телебиржи и телемагазины и т. д. Исчезни сегодня сети связи — и воцарит хаос. А ведь проводные линии связи так просто разрушить.

Наконец, в конце XX — начале XXI века человечество начало вырываться из плена проводов. Уровень развития микроэлектроники позволил выпускать массовые дешевые средства беспроводной связи. Бум сотовой связи, сравнимый разве что с ростом производства персональных компьютеров, не замедляется вот уже четверть века. Мобильных телефонов во всем мире уже больше, чем обычных проводных телефонных аппаратов, и это только начало. Фантастическими темпами развиваются технологии беспроводных локальных сетей, их догоняют персональные беспроводные сети и сети регионального масштаба. Видимо, все возвращается на круги своя — человечество тысячелетиями жило без проводов, хочет без них жить и впредь.

Бурное развитие беспроводных сетей передачи информации в России и во всем мире, о котором многие говорят как о беспроводной революции в области передачи информации [13, 100, 193], связано с такими их достоинствами, как:

- гибкость архитектуры, т. е. возможности динамического изменения топологии сети при подключении, передвижении и отключении мобильных пользователей без значительных потерь времени;
- высокая скорость передачи информации (1–10 Мбит/с и выше);
- быстрота проектирования и развертывания;
- высокая степень защиты от несанкционированного доступа;
- отказ от дорогостоящей и не всегда возможной прокладки или аренды оптоволоконного или медного кабеля.

Отметим, что современные телекоммуникационные технологии базируются на совокупности научных, технических и технологических достижений во многих областях, от микроэлектроники и схемотехники до теории связи, вычислительной техники и современных методов организации производства. Теория Максвелла оставалась мало кому понятной абстракцией до ее подтверждения опытами Герца. Кодовое разделение каналов и связь посредством шумоподобных сигналов не вышли бы из стен лабораторий и сложнейших военных систем, если бы не массовое появление дешевых процессоров цифровой обработки сигнала. Интернет остался бы ARPAnet'ом, если бы не лавиноподобное распространение персональных компьютеров и модемов. Сотовые телефоны и пейджеры, Wi-Fi-карты и цифровое телевидение никогда не увидели бы свет без интеграции успехов в самых разных областях — технических, законодательных, организационных, научных и т. д. Поэтому, описывая в настоящей монографии современные технологии беспроводных систем передачи информации (БСПИ), мы затронем не только вопросы теории передачи информации, принципов действия, архитектуры и методов анализа и проектирования беспроводных сетей, но и практическую реализацию аппаратуры беспроводных сетей. Кроме того, говоря о современных БСПИ, правомерно рассматривать как системы, изначально ориентированные на передачу данных (например, беспроводные локальные сети), так и телефонные системы (DECT, сотовая телефония), системы телевизионного и радиовещания и т. д.

В научно-технической литературе беспроводным технологиям уделяется огромное внимание. Регулярно проводятся специализированные международные конференции: International Symposium on Wireless Personal Multimedia Communication [214, 215]; IEEE International Conference on Mobile Ad-hoc and Sensor Systems [174]; International IFIP Networking Conference [209] и т. д.; имеется и целый ряд монографий по отдельным направлениям беспроводных сетей [9, 14, 93, 125, 131–133, 149, 153, 182, 183, 202].

Однако отсутствуют монографии, отечественные и зарубежные, которые описывали бы все современные и перспективные беспроводные сети, начиная с теоретических основ передачи информации, описания логической и физической структуры, методов проектирования до технической реализации беспроводных сетей передачи информации. В настоящей монографии делается попытка закрыть этот пробел.

В главе 1 монографии приведен обзор современных систем беспроводной связи. Рассмотрена классификация систем, приведены основные системы стандартов в этой области, включая модель взаимодействия открытых систем. Кратко рассматриваются основные методы доступа к среде в беспроводных сетях.

Глава 2 посвящена описанию технологий и архитектуры беспроводных сетей. Последовательно рассматриваются персональные беспроводные сети (IEEE 802.15.1 (Bluetooth) [65], IEEE 802.15.3, 3а и 4 [108]), локальные беспроводные сети передачи данных (IEEE 802.11 [105, 106] и DECT [98]), региональные (стандарт широкополосного вещания IEEE 802.16), сотовые [101, 102] и транкинговые сети, системы цифрового теле- и радиовещания [104], спутниковые системы и оптические атмосферные каналы связи [20, 50]. Изложена технология сверхширокополосной связи [103]. Весь материал представлен в исторической ретроспективе, что показывает как устойчиво сохраняющиеся тенденции, так и динамику, перспективы и основные направления развития систем беспроводной связи. Для ряда особо значимых беспроводных технологий приведены примеры аппаратной реализации устройств, описана их электронная компонентная база.

В главах 3–5 рассмотрены теоретико-информационные аспекты создания широкополосных беспроводных сетей. Приводятся основополагающие в теории информации теоремы Шеннона, Котельникова и Найквиста (глава 3). Рассмотрены методы сжатия и шифрования информации. Подробно описывается теория корректирующих кодов от простейших линейных и циклических кодов до последних модификаций низкоплотностных и турбокодов (глава 4). Рассматриваются основные системы модуляции и на их основе сигнально-кодовые конструкции для гауссова канала. В основу рассмотрения положен принцип обменных соотношений между скоростью передачи и помехоустойчивостью. Описаны сигнально-кодовые конструкции для каналов с межсимвольной интерференцией и переменных во времени каналов, построенных на основе ортогональной частотной модуляции (OFDM). Рассмотрены вопросы построения конструкций, методы передачи и приема, а также оценка качества каналов связи (глава 5).

Глава 6 посвящена описанию логической и физической структуры локальных беспроводных сетей, функционирующих под управлением протоколов IEEE 802.11, оценке их производительности. Рассматривается топологическая структура, методы кодирования, модуляция и защита информации в локальных сетях со стандар-

том IEEE 802.11, проведен сравнительный анализ стандартов IEEE 802.11a, b, g. Описываются аппаратура и фирмы-разработчики новейших типов чипсетов. Производительность локальной сети с протоколом IEEE 802.11 оценивается сначала для случая идеального канала, когда источниками потери производительности являются только коллизии станций, и высокой нагрузки на все станции (ситуации насыщения сети) [212]. В этом случае подробно исследуется эффективность двух механизмов передачи пакетов: механизма базового доступа и альтернативного механизма RTS-CTS. Анализируется влияние помех на производительность сети и эффективность механизма фрагментации пакетов; предлагается и изучается новый механизм отсрочки передачи [22, 173, 174, 209]. Исследуется также эффективность работы такой сети в условиях таких популярных приложений, как передача широковещательных сообщений и сети типа «Hot spot».

В главе 7 исследуется работа сетей городского и регионального масштаба, основой которых является протокол IEEE 802.11. Основными отличиями таких сетей от локальных являются скрытость оконечных станций друг от друга, что неизбежно приводит к высокой вероятности коллизий, а также высокая интенсивность помех. Разрабатываются аналитические и имитационные модели таких сетей в условиях распределенного управления, являющегося основным режимом работы протокола IEEE 802.11; исследуется эффективность механизма фрагментации пакетов в условиях городских и региональных сетей [21, 24, 25]. Особое внимание в главе (раздел 7.4) уделено оценке максимальной производительности, обеспечиваемой такими сетями при доступе в Интернет (т. е. при взаимодействии со стеком протоколов TCP/IP) [23, 211]. Данные аналитические (а также экспериментальные — см. раздел 7.7) исследования привели к выводу о неэффективности распределенного управления в сетях городского и регионального масштаба и необходимости использования альтернативного централизованного режима, который подробно аналитически исследуется и оптимизируется (см. также [3, 23, 210, 213]) в разделах 7.5 и 7.6. В частности, предлагается и исследуется обобщенная адаптивная динамическая стратегия отложенного опроса оконечных станций, позволяющая избежать потерь, связанных с бесполезным опросом станций, не имеющих готовых пакетов для передачи [214].

Раздел 7.7 посвящен экспериментальным исследованиям и оптимизации Московской городской сотовой сети Radionet, соты которой работают под управлением протокола IEEE 802.11. В этом разделе проведен анализ особенностей беспроводного трафика, проходящего через данную сеть, динамики изменения интенсивности помех и с учетом данных особенностей разработан метод оптимизации фрагментации пакетов [56, 57]. Наконец, в заключительном разделе главы описывается и решается задача синтеза топологической структуры региональной сети, включающая оптимальное размещение базовых станций и прикрепление к ним локальных сетей абонентов при наличии ограничений на электромагнитную совместимость.

В главе 8 рассматривается перспективный стандарт региональных беспроводных сетей IEEE 802.16-2004. Приведено описание радиointерфейсов, методов модуляции и доступа к каналам, системы управления потоками, механизмов уровня управления доступом к каналу передачи (MAC-уровня) [107, 109, 110]. Рассматриваются особенности реализации аппаратуры и сетей стандарта IEEE 802.16.

Глава 9 посвящена описанию архитектуры и технических средств беспроводных региональных сетей. Рассматривается архитектура сети Radionet и логическая и аппаратная структура отечественного радиомаршрутизатора «РАПИРА». Рассмотрены основные принципы построения современных антенных систем широкополосной связи. Основное внимание уделено обобщению новых технических решений в этой области, в частности детально изложены последние достижения теории и техники фрактальных антенных систем. Впервые в отечественной литературе подробно рассмотрены основы схемотехники и принципы функционирования smart-антенн и базирующейся на них технологии ММО, используемой в новых широкополосных стандартах IEEE 802.11 и IEEE 802.16-2004.

Дано также описание оригинальной технологии построения широкополосных беспроводных сетей с использованием высотных платформ. Приводится классификация высотных платформ, включая использование привязных аэростатов и высотных дирижаблей, располагающихся на высоте 22 км. Описывается опыт разработки и реализации отечественного проекта широкополосной сети «БАРС» на базе привязных аэростатов [210, 215].

Авторы приносят благодарность Г.Р. Тарловскому, написавшему разделы 3.4–3.6, В.И. Слюсару, написавшему разделы 9.4–9.9, И.Н. Астафьевой и другим сотрудникам отдела сетей передачи информации Института проблем передачи информации РАН, принимавшим участие в подготовке книги. Особая признательность выражается академикам Ю.В. Гуляеву и Н.А. Кузнецову за поддержку и ценные советы при написании книги.

ГЛАВА 1

Беспроводные сети передачи информации. Основные понятия

1.1 Исторический очерк развития сетевых технологий

Беспроводные сети передачи информации, как следует из их названия, базируются на совокупности двух групп технологий — беспроводной передачи информации и сетевого взаимодействия. Исторически эти технологии зародились еще в позапрошлом веке. Родоначальником всех электронных сетей (систем) передачи данных, видимо, следует считать американского художника Самуэля Финли Бриза Морзе. В 1837 году он разработал свою систему электросвязи по металлическому проводу и дал ей название «телеграф». Годом позже он дополнил ее знаменитой азбукой Морзе, т. е. механизмом кодирования источника, обязательным элементом всех современных сетей. 24 мая 1844 года между Балтимором и Вашингтоном состоялся первый публичный сеанс телеграфной связи. Уже через 14 лет был проложен первый трансатлантический кабель — правда, просуществовал он лишь 26 дней.

В 1874 году французский инженер Жан Морис Эмиль Бодо (Baudot) изобрел телеграфный мультиплексор, позволявший по одному проводу передавать до шести телеграфных каналов. Значимость этого изобретения и авторитет Бодо были столь высоки, что, когда в 1877 году другой французский инженер Томас Муррэй разработал первый в истории символьный телеграфный код с фиксированным размером символа (5 бит на символ), он назвал его кодом Бодо. Известный также под названием телексный код, он с незначительными изменениями применяется и сегодня (наиболее распространенная версия — стандартизированный Международным консультативным комитетом по телефонии и телеграфии (ССТТ) Международный алфавит № 2). В честь Бодо названа и единица измерения скорости передачи телекоммуникационных символов (бод).

Следующий шаг сделали изобретатели телефона — профессор физиологии органов речи Бостонского университета Александр Грэйхем Белл при участии Томаса Ватсона (1875 год, приоритет от 14 февраля 1876 года) и независимо от

них — Элайша Грей в Чикаго. Последнему также принадлежит немалая роль в развитии сетевых технологий. Именно он в 1888 году запатентовал *Telautograph* — первое устройство передачи факсимильных сообщений. Но это были лишь предпосылки сетей, а именно способы формирования канала связи и работы в нем. Сеть — это совокупность многих каналов, которыми необходимо управлять. В первых сетях начиная с 1880 года этим занимались телефонистки (вернее телефонисты) методом установки птекеров в коммутационном поле.

С 1889 года начался новый этап в развитии сетевых технологий — владелец бюро похоронных услуг из Канзас-Сити Элмон Браун Строуджер разработал систему автоматической коммутации каналов. Именно ему принадлежит приоритет в создании шагового искателя и декадно-шаговых АТС. Предание гласит, что Строуджер столкнулся с промышленной диверсией — жена его конкурента по цеху в Канзас-Сити работала телефонисткой и все звонки гробовщику направляла своему мужу — видимо, это один из первых случаев электронного шпионажа. Изобретение Строуджера оказалось столь удачным, что в 1891 году он основал компанию *Strowger Automatic Exchange* (с 1901 года — *Automatic Electric*, сегодня — отделение компании *General Telephone and Electronics* — GTE). Первая АТС этой компании емкостью 99 номеров была запущена в коммерческую эксплуатацию в 1892 году (г. Ла-Порт, шт. Индиана). Примечательно, что на первых телефонных аппаратах для работы с АТС номер набирался посредством кнопок. В 1897 году компания Строуджера представила прототип первого аппарата с дисковым номеронабирателем.

В 1885 году произошло еще одно ключевое для сетевых технологий событие. Первые АТС обеспечивали одновременное соединение всех возможных пар абонентов. Очевидно, что при росте номерной емкости коммутационные матрицы становились невероятно дорогими и сложными. Впервые возникла проблема доступа к ограниченному коммутационному ресурсу. Ее разрешил российский инженер М.Ф. Фрейденберг, показавший, что для 10 тыс. абонентов достаточно обеспечить возможность одновременного соединения любых 500 пар. Отметим, что результат Фрейденберга справедлив и сегодня, для современных АТС: на 10 тыс. номеров допустимая вероятность предоставления соединения составляет 0,125. В 1895 году М.Ф. Фрейденберг совместно с другим русским инженером С.М. Бердичевским-Апостоловым разработали и запатентовали в Великобритании АТС с так называемым предыскателем, выбиравшим свободный комплект линейных искателей при снятии абонентом трубки. Предыскатель и его принцип свободного поиска стал основой для проектирования всех будущих АТС. Примерно с 1910 года (к окончанию срока действия патента Строуджера) началось массовое внедрение электромеханических АТС. Работу, начатую М.Ф. Фрейденбергом, до логического завершения довел датский математик А.К. Эрланг, опубликовавший в 1909 году ставшую классической работу «*The Theory of Probabilities and Telephone Conversations*» («Теория вероятностей и телефонные переговоры»), в которой предложил формулы для вычисления числа абонентов АТС, желающих одновременно вести разговоры.

Работы А.К. Эрланга положили начало нового научного направления — теории очередей (теории массового обслуживания), широко используемой первоначально для расчетов в телефонии, а затем при проектировании сетей передачи

информации. Значительный вклад в развитие теории очередей внес выдающийся российский математик Александр Яковлевич Хинчин (Математическая теория стационарной очереди// Математический сборник. Т. 39, № 4. 1932; О формулах Эрланга в теории массового обслуживания// Теория вероятностей и ее применения. Т. 7. Вып. 3. 1962), выполнивший ряд оригинальных исследований для Московской телефонной сети.

В том же 1909 году генерал-майор корпуса связи США доктор философии Джордж Оуэн Скварер изобрел метод посылки по телефонной линии нескольких радиogramм одновременно — родился метод частотного разделения каналов.

В 1928 году американский физик-электрик и изобретатель Гарри Найквист в статье «Некоторые вопросы теории телеграфной передачи» («Certain Topics in Telegraph Transmission Theory») изложил принципы преобразования аналоговых сигналов в цифровые — родилась знаменитая теорема Найквиста. В СССР ее называли теоремой Котельникова, хотя Владимир Александрович опубликовал ее через пять лет после Найквиста. Но история все нивелирует — основополагающая теорема Клода Элвуда Шеннона о пропускной способности канала (1948 год) была сформулирована Котельниковым в его докторской диссертации годом раньше, в 1947 году. Однако у нас ее называют теоремой Шеннона.

В 1938 году американец А.Х. Риверс патентует метод преобразования сигнала из аналоговой формы в цифровую для коммутации и передачи, названный импульсно-кодовой модуляцией (ИКМ). Этот метод впервые был практически реализован учеными из Bell Laboratories Клодом Шенноном, Джоном Р. Пирсом и Бернардом М. Оливером в быстродействующей цифровой передающей системе, позволившей транслировать несколько телефонных разговоров по одному каналу с высоким качеством, — появилась система с временным разделением (уплотнением) каналов.

Начиная с 50-х годов сетевые и беспроводные технологии начали сближаться настолько тесно, что зачастую грань между ними провести уже трудно.

Беспроводные технологии также зарождались в XIX веке. Идея носилась в воздухе, вплотную к ней подошли такие ученые, как Г. Герц, О. Лодж, Э. Бранли. В 1892 году английский ученый Вильям Крукс теоретически показал возможность и описал принципы радиосвязи. В 1893 году сербский ученый Никола Тесла в США продемонстрировал передачу сигналов на расстояние. Тогда это событие не вызвало должного резонанса, возможно, потому, что Н. Тесла, работы которого существенно опережали время, интересовался беспроводной передачей на расстояние не информации, а энергии.

С 1878 года над проблемой беспроводной связи работал преподаватель Минных классов в Кронштадте Александр Степанович Попов. В 1884 году он изобрел первую приемную антенну, создал прибор для регистрации грозových разрядов на основе когерера — стеклянной трубки, заполненной металлическими опилками. Под воздействием электромагнитного поля проводимость этой трубки резко возрастала. 7 мая 1895 года на заседании физического отделения Российского физико-химического общества состоялся его исторический доклад «Об отношении металлических порошков к электрическим колебаниям». Тогда А.С. Попов продемонстрировал свой прибор для регистрации грозových разрядов («грозоотметчик») и высказал мысль о возможности его применения для беспроводной

связи. Первая публичная демонстрация прототипа всех грядущих беспроводных систем состоялась 24 марта 1896 года на заседании того же физико-химического общества. А.С. Попов передал на расстояние 250 м, возможно, первую в мире радиограмму, состоявшую из двух слов «Генрих Герц».

С 1894 года успешно экспериментировал физическими приборами для генерации и регистрации электромагнитных колебаний и двадцатилетний итальянский юноша Гульельмо Маркони, будущий Нобелевский лауреат. В 1895 году он установил связь на расстоянии порядка двух миль, в 1896 году запатентовал свое изобретение (в 1943 году его патенты были аннулированы в пользу Н. Тесла [220]), в 1901-м установил радиосвязь через Атлантику.

В 1906 году Ли де Форест создал первую электронную лампу (триод) — появилась возможность строить электронные усилители сигналов. С тех пор беспроводная связь развивалась — и продолжает по сей день — семимильными шагами, главным образом благодаря достижениям электроники. Отметим лишь основные вехи.

С 20-х годов началось коммерческое радиовещание (посредством амплитудной модуляции). В 1933 году Эдвин Ховард Армстронг изобрел частотную модуляцию (ЧМ), с 1936 года началось коммерческое ЧМ-радиовещание. В 1946 году компании AT&T и Bell System приступили к эксплуатации системы подвижной телефонной связи (MTS) для абонентов с автомобильными радиотелефонами (20 Вт). Для полудуплексной связи использовалось 6 каналов шириной по 60 кГц на частоте 150 МГц, однако из-за межканальной интерференции число каналов вскоре сократили до трех. Система позволяла соединяться с городской телефонной сетью.

12 августа 1960 года был выведен на орбиту высотой 1500 км первый спутник связи — американский космический аппарат (КА) «Echo-1» (Эхо-1). Это был надувной шар с металлизированной оболочкой диаметром 30 м, выполнявший функции пассивного ретранслятора. Через два года, 10 июля и 13 декабря 1962 года, в США на низкие орбиты были запущены соответственно КА «Telstar I» и «Relay-1» — первые спутники с активными ретрансляторами. Мощность их передатчиков не превышала 2 Вт. 19 августа 1964 года впервые спутник связи был выведен на геостационарную орбиту. Это был также американский «Syncom-3» (первые две попытки вывода в 1963 году были неудачными). На следующий день был создан международный консорциум спутниковой связи Intelsat (International Telecommunications Satellite Organization), который стал крупнейшей международной организацией в области спутниковой связи. Сегодня ее услугами пользуются более чем в 200 странах, причем в начале 2001 года 2/3 всего международного трафика передавалось через спутники Intelsat. 23 апреля 1965 года был выведен на орбиту и начал успешно работать первый отечественный спутник связи «Молния-1» (также с третьей попытки). Началась эра спутниковой связи.

В истории сетевых технологий очередной этап начался в 60-е годы прошлого столетия и связан с массовым появлением компьютеров. Возникла потребность в передаче большого объема данных, зародилось понятие локальной вычислительной сети (ЛВС). Был разработан механизм коммутации сообщений (пакетов). В 60-е годы над построением сети с коммутацией пакетов работали (параллельно, практически ничего не зная друг о друге) специалисты в трех организациях: в

Массачусетском технологическом институте (MIT), в корпорации RAND (центр стратегических исследований ВВС США), в Национальной британской физической лаборатории (NPL). Пионерской работой в этой области явилась диссертация Леонарда Клейнрока на соискание степени доктора философии в MIT «Информационный поток в больших коммуникационных сетях» («Information Flow in Large Communication Nets», 1961 год). В 1964 году была опубликована работа сотрудника корпорации RAND Пола Барана «О распределенных коммуникациях» («On Distributed Communications»). В ней были сформулированы принципы избыточной коммуникативности и показаны различные модели формирования коммуникационной системы, способной успешно функционировать при наличии значительных повреждений. В 1965 году Лоуренс Робертс из MIT совместно с Томасом Меррилом связал компьютер TX-2 в Массачусетсе с ЭВМ Q-32 в Калифорнии по низкоскоростной коммутируемой телефонной линии. Так была создана первая нелокальная компьютерная сеть. Она убедительно продемонстрировала, что сеть с коммутацией соединений (каналов) неприемлема для построения ЛВС.

В 1962 году в журнале «Коммунист» (№ 12) появилась статья академика АН СССР Александра Александровича Харкевича «Информация и техника». В ней впервые в мире были сформулированы основные принципы создания единой сети связи (ЕСС), предугадана важность цифровых методов передачи и коммутации различных видов информации в цифровой форме. ЕСС, по мнению А.А. Харкевича, должна представлять собой крупнейший инженерный комплекс, объединяющий всю существующую сеть связи и развивающийся путем планомерного ее наращивания в органическом взаимодействии с системой вычислительных, управляющих и справочных центров.

Знаковыми для сетевых технологий стали 1967–1968 годы. В NPL заработала первая ЛВС с пакетной коммутацией, во многом благодаря ее директору Дональду Дэвису. Сеть работала с пиковой скоростью до 768 кбит/с (в начале 70-х она объединяла порядка 200 компьютеров со скоростью обмена до 250 кбит/с). В том же 1968-м сотрудник шведского отделения компании IBM Олаф Содерблум разработал сеть Token Ring. МО США одобрило версию первого в мире стандарта на ЛВС — MIL-STD-1553 (протокол обмена данными по общему последовательному каналу посредством манчестерского линейного кода с выделенным контроллером; отечественный аналог — ГОСТ 26765.52–87 в бортовых системах применяется до сих пор).

Но самое главное, в октябре 1967 года был представлен начальный план сети ARPANET, развитием которой занимался департамент методов обработки информации IPTO (Information Processing Techniques Office) агентства перспективных исследовательских проектов ARPA (Advanced Research Projects Agency) МО США. В декабре 1968 года группа во главе с Фрэнком Хартом из компании Bolt, Beranek и Newman (BBN) выиграла конкурс ARPA на создание так называемого интерфейсного процессора сообщений (Interface Message Processor). В 1969 году в рамках программы ARPANET в Калифорнийском университете в Лос-Анжелесе «отец» пакетной коммутации Леонард Клейнрок построил первый узел ARPANET — прообраз грядущего Интернета. В том же году компания BBN установила в Калифорнийском университете первый интерфейсный процессор сообщений и подключила к нему первый компьютер. Второй узел был

образован в Стэнфордском исследовательском институте (SRI). Двумя следующими узлами ARPANET стали Калифорнийский университет в Санта-Барбара и университет штата Юта. Эмбрион Интернета начал делиться.

В 1970 году появилась первая пакетная радиосеть передачи данных (через спутник) — знаменитая ALOHA (aloha — приветствие в гавайском диалекте английского языка). Ее разработал и построил Норман Абрамсон (совместно с Франком Куо и Ричардом Биндером) из Гавайского университета. Сеть связывала различные университетские учреждения, разбросанные по отдельным островам Гавайского архипелага. В 1972 году ALOHA связали с сетью ARPANET. В ALOHA был реализован принцип подтверждения и повторной отправки пакетов (ARQ), а также механизм множественного доступа к каналу с контролем несущей CSMA. Тогда же начали развиваться проекты создания пакетных радиосетей, в том числе спутниковых.

В октябре 1972 года известный специалист из компании BBN Роберт Кан на международной конференции по компьютерным коммуникациям впервые публично продемонстрировал работу сети ARPANET. В 1974 году появляется статья Вирта Серфа (сотрудника Стэнфордского исследовательского института) и Роберта Кана (V.G.Cerf, R.E. Kahn. A protocol for packet network interconnection// IEEE Trans. Comm. Tech. Vol. COM-22. V 5. May 1974. P. 627-641), в которой впервые была описана концепция протокола TCP/IP. В том же году компания BBN запустила первую открытую службу пакетной передачи данных (коммерческая версия ARPANET) — известный сегодня любому специалисту Telnet.

В 1973 году сотрудник исследовательского центра компании Xerox в Пал-Альто Роберт Метклаф, до прихода в Xerox защитивший в MIT докторскую диссертацию в области теории пакетной передачи информации и участвовавший в создании сети ARPANET, представил своему руководству докладную записку, в которой впервые появилось слово Ethernet (эфирная сеть). В том же году Метклаф совместно с Дэвидом Боггсом построил первую Ethernet-ЛВС, связывавшую два компьютера со скоростью 2,944 Мбит/с. В основу технологии Ethernet был положен усовершенствованный принцип CSMA/CD с обнаружением коллизий. Через шесть лет, в 1979 году, при активном участии Б. Метклафа три ведущие в своих областях компании США (Xerox, Intel и Digital Equipment (DEC)) начали процесс стандартизации протокола Ethernet, успешно завершившийся через год. В том же 1979 году Метклаф при участии DEC основал знаменитую компанию 3COM для выпуска Ethernet-совместимого оборудования.

В 1976 году CCITT выпустила рекомендацию X.25, которая стала первым и чрезвычайно успешным стандартом сети с пакетной передачей данных по выделенному каналу (Interface between DTE and DCE for Terminal Operations in Packet Mode and Connected to Public Data Networks by Dedicated Circuit). Массовая пакетная коммуникация стала реальностью.

В 1977 году будущий вице-президент компании Sony Марио Токорои и другой японский ученый Киичироу Тамару предложили метод адаптации технологии Ethernet к передаче данных через радиоканал посредством механизма подтверждений (Acknowledging Ethernet). Эта работа заложила основу будущих беспроводных ЛВС (IEEE 802.11 и IEEE 802.15).

В 1978 году в Бахрейне телефонная компания Batelco (Bahrain Telephone Company) впервые в мире начала эксплуатацию коммерческой системы беспроводной телефонной связи, которая считается первой в мире реальной системой сотовой связи. Две зоны с 20 каналами в диапазоне 400 МГц обслуживали 250 абонентов. Использовалось оборудование японской компании Matsushita Electric Industrial. В том же году в Чикаго компания AT&T начала испытания сотовой системы Advanced Mobile Phone Service (AMPS), работающей в диапазоне 800 МГц. Сеть из 10 зон охватывала связью 54 тыс. км².

В 1977 году Деннис Хайес основал компанию Hayes Microcomputer Products и выпустил на рынок первый массовый модем Micromodem II для персональных компьютеров (Apple II). Он работал со скоростью 110/300 бит/с и стоил 280 долл. В 1979 году в Женеве CCITT утверждает первую модемную рекомендацию V.21, определяющую стандартный протокол модуляции на скорости 300 бит/с.

В 1978 году международная организация по стандартизации ISO утвердила семиуровневую модель взаимодействия открытых систем.

Новый этап начался в 1980 году, когда протокол TCP/IP был принят в качестве военного стандарта США. Годом раньше пакетная радиосеть заработала в военной базе США Форт Брэгг. Через три года, в 1983 году, сеть ARPANET была переведена на протокол TCP/IP (взамен действовавшего изначально NCP). Произошло выделение из ARPANET (которую вскоре все стали называть Интернетом) сети MILNET, обслуживающей оперативные нужды МО США.

События периода 60-х годов в области сетевых технологий описаны во множестве книг, воспроизводить которые здесь невозможно да и не нужно. За каждой датой, за каждым событием стоят напряженная работа и выдающиеся достижения специалистов всего мира. В это время сетевые технологии непрерывно развивались в сторону повышения быстродействия и надежности сетей передачи информации, возможности интегрированной передачи данных, голоса и видеоинформации. Так, в области локальных сетей было создано семейство технологий Ethernet-Fast Ethernet-Gigabit Ethernet, обеспечивающих иерархию скоростей 10/100/1000 Мбит/с. В глобальных сетях произошел переход от технологии X.25 к технологии Frame Relay, использованию стека протоколов TCP/IP, ATM и Gigabit Ethernet.

Чтобы завершить краткий исторический обзор, отметим, что и в СССР работало немало выдающихся ученых и специалистов в области систем связи, в том числе и беспроводной. Уже в 70–80-е годы проектировались и строились современные сети связи. Назовем, например, систему цифровой телефонной связи «Кавказ-5», многочисленные ведомственные сети связи. Хорошо известны системы «Сирена» (первая в СССР гражданская сеть пакетной коммутации) и «Экспресс» для автоматизации бронирования и продажи авиа- и железнодорожных билетов соответственно. Но, видимо, закрытость как самих работ, так и общества никак не согласовывались с концепцией открытых сетей. Возможно, именно поэтому изначально созданная на деньги МО США открытая сеть Интернет завоевала весь мир, породила множество сетевых технологий, стимулировала развитие смежных отраслей, прежде всего разработку соответствующей аппаратуры и элементной базы для нее, т. е. микроэлектронику.

Возможно, именно Интернету мы исторически обязаны тем, что сегодня широкополосные беспроводные сети получили столь бурное развитие. Их появление было бы невозможно без соответствующей полупроводниковой элементной базы. А она, в свою очередь, не может появиться, если нет массового (многомиллионного) спроса. Историческая заслуга и гениальное провидение тех, кто в 60-е годы начинал работы по сетям пакетной передачи, в том, что они изначально сумели сформулировать принципы будущей глобальной сети и воплотили их. Тем самым был создан рынок устройств для работы в сети, ставший основой для промышленности и науки в этой области. Не случайно первым директором (с 1962 года) департамента IPTO в ARPA, т. е. человеком, руководившим финансированием научных исследований в области компьютерных сетей, был Джозеф Карл Ликлайдер из Массачусетского технологического института. Еще в начале 60-х он сумел предвидеть появление глобальной сети взаимосвязанных компьютеров. Ему принадлежит ряд публикаций о концепции «галактической сети» (J.C.R. Licklider. On-Line Man Computer Communication, August 1962).

Разумеется, не менее основополагающим для беспроводных сетей стало массовое появление персональных компьютеров и развитие сотовой телефонии, а также стремительное развитие полупроводниковых технологий (создание дешевых сигнальных процессоров и микроконтроллеров, аналоговых СВЧ интегральных схем). Вот эти слагаемые и привели сегодня к тому, что высокоскоростные беспроводные СПИ готовы избавить мир от пут проводных коммуникаций.

1.2 Классификация и технологии беспроводных сетей

Классификация чего бы то ни было — задача неблагодарная, поскольку и критериев классификации можно разработать достаточно много, и реальные объекты могут не укладываться в четкие границы определенного класса, да и по мере развития устоявшиеся системы классификации могут устаревать. Все это справедливо и для беспроводных сетей передачи информации (БСПИ). Поэтому остановимся на наиболее популярных способах ранжирования различных беспроводных систем. Обычно БСПИ подразделяют:

- по способу обработки первичной информации — на цифровые и аналоговые;
- по ширине полосы передачи — на узкополосные, широкополосные и сверхширокополосные;
- по локализации абонентов — на подвижные и фиксированные;
- по географической протяженности — на персональные, локальные, региональные (городские) и глобальные;
- по виду передаваемой информации — на системы передачи речи, видеoinформации и данных.

Вполне справедливы и системы градации на основе используемой технологии (спутниковые сети, атмосферные оптические линии и т. п.), по назначению и др.

Поскольку в центре нашего внимания находятся цифровые беспроводные широкополосные системы, приведем их отличительные признаки, охарактеризовав и «сопредельные» системы. Термин «беспроводность» определяется легко — отсутствует соединительный провод (оптоволоконный или медный кабель). Также относительно просто определить, цифровая система или нет. К цифровым относят системы, у которых входная аналоговая информация (например, голос, аналоговый телевизионный сигнал и т. п.) первоначально преобразуется в цифровую (дискретную) форму. Однако уже здесь возникает некоторая нечеткость. В самом деле, любой сигнал при передаче через физический канал имеет чисто аналоговый вид, он в принципе не должен быть дискретным (чем дальше форма сигнала от бесконечной синусоиды, тем больше паразитных гармоник и связанных с ними неприятностей), чего добиваются специальными методами. Поэтому термин «цифровая система» говорит только о том, что в ней входящие аналоговые данные оцифрованы и обрабатываются (фильтрация, скремблирование, коммутация) преимущественно цифровыми методами.

Еще сложнее с шириной полосы. Строгого определения тут нет. Обычно полагают, что если ширина спектральной полосы F , в которой работает система, много меньше центральной частоты этой полосы f_c , то система узкополосная (т. е. $F/f_c \ll 1$). В противном случае система широкополосная. Критерий весьма расплывчат. В области цифровых систем передачи приводят и другие определения широкополосности [93]: например, система широкополосная, если передаточная функция канала в этой полосе существенно изменяется в зависимости от частоты (т. е. передаточная функция в рабочей полосе узкополосной системы практически не зависит от частоты). Очевидно, что определения эти достаточно расплывчатые. В нашем случае под термином «широкополосная система» мы будем понимать такие системы, где проявляются специфические эффекты и свойства, связанные с широкой рабочей полосой. Поэтому точный критерий не существует да и не возможен.

Подразделение на мобильные и подвижные системы, казалось бы, столь простое, на самом деле также не является тривиальным. Следует различать собственно возможность мобильности абонентов, предоставляемую технологией, и подразделение на мобильную и фиксированную службы связи, связанное с вопросами частотного распределения и лицензирования. Наиболее характерным примером такой двухсмысленности является история появления в России беспроводной телефонной связи стандарта IS-95 (CDMA). Оборудование этого стандарта изначально было разрешено к использованию в нашей стране только для предоставления услуг фиксированной связи. Однако, как известно, IS-95 является стандартом мобильной сотовой связи. Технологически его никак нельзя «закрепить». Аналогичная неопределенность сложилась сейчас и в спутниковой связи. Если же говорить с технической точки зрения, ограничивать мобильность может чувствительность технологии связи к скорости движения абонента, сложность перехода из одной зоны обслуживания в сопредельную без разрыва связи, восприимчивость к кратковременным пропаданиям связи и т. п.

Подразделение по размеру зоны обслуживания также достаточно условно, если рассматривать соседние градации. К персональным сетям (WPAN — wireless personal area network) относят системы с радиусом действия от сантиметров до

нескольких метров (до 10–15 м). Основное назначение таких сетей состоит в замещении кабельной системы для связи оборудования (например, компьютера и периферийных устройств). При этом мощность излучения передатчиков, как правило, не превышает 10 мВт. Локальные сети (WLAN — wireless local area network) подразумевают взаимную удаленность устройств на расстоянии до сотен метров и мощности передатчиков до 100 мВт. Это сети, предназначенные для объединения устройств в пределах локальной зоны (здания, предприятия и т. п.). Отметим, что на основе стандартов локальных беспроводных сетей вполне успешно строят и сети городского масштаба. Например, применение в этом качестве используют такие технологии, как DECT и IEEE 802.11.

К сетям городского масштаба (региональным) можно отнести множество различных технологий. Это и наземное теле- и радиовещание, и сотовая связь, и транковые системы. Недавно появилось семейство стандартов на широкополосные беспроводные сети городского масштаба IEEE 802.16. Если же говорить о глобальных беспроводных системах передачи данных, то они представлены спутниковыми системами связи. Однако с учетом того, что, например, практически все сети сотовой телефонии так или иначе связаны друг с другом, все они разрабатываются с учетом возможности взаимодействия, можно (правда, с некоторой натяжкой) говорить и о глобальных сотовых сетях.

Особой градацией является подразделение в зависимости от типа передаваемой информации, например, на системы передачи речи (или видеoinформации) и несинхронных данных. С одной стороны, речь — это один из видов информации. После оцифровки поток речевых данных по виду не отличим от потока любой другой информации. Развитие цифровых технологий в различных областях телекоммуникаций (например, в проводной телефонии) давно продемонстрировало эффективность цифровых методов обработки, когда и речь, и данные обрабатываются едиными способами. С другой стороны, потребность в информации разного вида уже сделала реальной интеграцию различных информационных сетей (телефония, телевидение, сети передачи цифровых данных, телеметрия) на бытовом уровне. По единому каналу передаются данные самой различной природы. Поэтому можно достаточно уверенно предположить, что недалек тот день, когда вся речевая информация будет обрабатываться исключительно цифровыми методами. Здесь можно было бы остановиться, но возникает важный нюанс. Каждому виду информации свойственны характерные требования при передаче. Человек чувствует задержку передачи речи, когда она превышает 0,25 с. При задержках около 0,5 с восприятие речи для многих становится неприемлемым. Причем дело не только собственно в задержке, но и в неизбежном при дуплексной связи эхосигнале, который при таких задержках устранить нереально. С другой стороны, речевая информация малочувствительна к спорадическим помехам и потерям данных. Это означает, что при пакетной передаче речи важно, чтобы задержки распространения сигнала в канале были минимальными, а маршрутизация и восстановление потока данных из пакетов (даже если их последовательность нарушена) происходили в реальном времени. При этом допустима даже потеря отдельных пакетов. Аналогична ситуация и с передачей видеoinформации — задержка между приемом отдельных пакетов (например, MPEG-2) не должна превышать некоего заданного значения, но потеря пакета, как правило, допустима. Совершенно иные требования предъявляются к передаче телеметрической

информации, текстовых данных и т. п. Здесь, как правило, не важен режим реального времени (в определенных пределах), но и недопустима потеря данных. Учет этих особенностей может приводить к созданию особых технологий, ориентированных на трансляцию определенных видов информации. Характерным примером является появление технологии Frame Relay — способа пакетной передачи, при котором не происходит проверок прохождения отдельных пакетов (в отличие от традиционных сетей пакетной коммутации X.25 с подтверждением и повторной передачей каждого пакета).

Приведенные выше рассуждения показывают, что любое определение, так или иначе ранжирующее БСПИ, не стоит воспринимать буквально и уж тем более не стоит удивляться применению той или иной технологии «не по назначению».

1.3 Стандартизация в области телекоммуникаций

Важнейшим аспектом развития современных телекоммуникационных систем является их стандартизация. Стандартизация необходима всем обитателям мира телекоммуникаций, включая производителей электронных компонентов, изготовителей аппаратуры, разработчиков сетей и конечных пользователей. Прежде всего стандартизация означает массовость производства, что ведет к низким ценам и широкому распространению технологии. Разумеется, выбор и утверждение стандарта — это процесс не только технический, но и политический. Как правило, различные фирмы зачастую прорабатывают альтернативные варианты будущей технологии. От того, какой из них будет утвержден в качестве стандарта, зависят и объемы будущих прибылей. Поэтому, чтобы стандарт действительно стал общепризнанным, стандартизирующая организация должна быть чрезвычайно авторитетной, а сама процедура утверждения — максимально открытой и беспристрастной.

«Головной» организацией в мире в области стандартизации в телекоммуникациях является Международный телекоммуникационный союз (ITU — International Telecommunications Union), работающий под эгидой ООН. После реорганизации 1 марта 1993 года два его сектора вобрали в себя три важнейшие стандартизирующие организации. В сектор радиосвязи (ITU-R, ITU Radiocommunication Sector) вошли Международный консультативный комитет по радиовещанию (CCIR — International Radio Consultative Committee) и Международный комитет по регистрации радиочастот (IFRB — International Frequency Registration Board). До этого ключевую роль в распределении спектра играли именно эти организации. Они также занимались вопросами спутниковых систем связи, глобальных радиосистем и др. Сектор телекоммуникаций международного телекоммуникационного союза ITU-T (ITU Telecommunication Standardization Sector) стал преемником Международного консультативного комитета по телерафии и телефонии (CCITT — Consultative Committee for International Telephone and Telegraphy), долгие годы издававшего свои знаменитые «цветные книги» — сборники стандартов в области телекоммуникаций. Третий сектор ITU — исследовательский

сектор ITU-D, созданный на базе организованного в 1989 году Бюро телекоммуникационных исследований (Telecommunication Development Bureau — TDB).

Надо отметить, что не случайно именно ITU играет главенствующую роль в области международных телекоммуникационных стандартов. История этой организации неразрывно связана с историей телекоммуникационной индустрии. Впервые аббревиатура ITU появилась 140 (!) лет назад, 17 мая 1865 года в Париже, когда представители 20 государств приняли первую международную телеграфную конвенцию и учредили Международный телеграфный союз — International Telegraph Union (ITU). Напомним, тогда телеграф был единственным видом электросвязи! С 1885 года ITU занимался и вопросами международной стандартизации телефонии. В 1906 году под эгидой ITU в Берлине прошла первая международная конференция по вопросам радиотелеграфии, на которой была принята первая конвенция, заложившая основу того, что сегодня называют частотным регулированием.

К середине 20-х годов были основаны Международные консультативные комитеты в области телефонии (The International Telephone Consultative Committee — CCIF, 1924 год), телеграфии (The International Telegraph Consultative Committee — CCIT, 1925 год) и радиовещания (CCIR, 1927 год). В 1932 году в Мадриде ITU обрел свое современное название — International Telecommunication Union (официальная дата переименования — 1 января 1934 года). 15 октября 1947 года в Атлантик-сити ITU получил статус агентства только что организованной ООН. В том же году был основан и Международный комитет по регистрации радиочастот IFRB. В 1956 году CCIT и CCIF слились в единую организацию CCITT. 1 марта 1993 года завершилась одна из крупнейших в практике реорганизаций, в результате чего ITU обрел свой современный вид. Отметим, что современная структура ITU как нельзя лучше отражает ситуацию в области современных телекоммуникаций: тесное переплетение и единение технологий проводных и беспроводных, аналоговых и цифровых.

Огромную роль в области утверждения международных стандартов играет Международная организация по стандартизации ISO (International Organization for Standardization). Это сеть институтов стандартизации 148 различных стран. Само название организации ISO не сокращение (было бы IOS), а производное от греческого слова isos (равный). Если ITU, как агентство ООН, представляет собой межправительственную организацию, то ISO — неправительственный орган. Среди членов ISO не только правительственные, но и частные организации. Прародителем ISO стала основанная в 1906 году и успешно действующая до сих пор Международная электротехническая комиссия (МЭК, IEC — International Electrotechnical Commission), занимающаяся вопросами стандартизации в области электротехники и электроники. В 1926 году была создана ISA (International Federation National Standardizing Associations) — международная федерация национальных стандартизирующих организаций. С началом Второй мировой войны ISA прекратила существование. Однако в 1946 году на конференции национальных стандартизирующих организаций, проходившей с 14 по 26 октября в Лондоне, делегаты 25 стран возродили ее, но под названием ISO (официальное начало работы 23 февраля 1947 года). Важную роль в рождении ISO сыграл созданный в 1944 году Координационный комитет стандартов Объединенных Наций (United

Nations Standards Coordinating Committee), влившийся в ISO. Сегодня ISO играет важнейшую роль в области стандартизации в телекоммуникациях, работая в плотной кооперации с IEC. Можно сказать, что IEC и ISO поделили сферы влияния: IEC — стандарты в области электроники и электротехники, ISO — все остальное. Они используют единую систему нумерации, и в кодах стандартов зачастую фигурирует название обеих организаций, например ISO/IEC 8802-3.

Отметим, что и ITU, и ISO, и IEC выступают скорее как самые авторитетные утверждающие организации. Сегодня общемировая практика такова, что собственно технической проработкой будущих стандартов занимаются ведущие национальные и межнациональные организации. Прежде всего это Европейский институт стандартизации в области телекоммуникаций ETSI (European Telecommunications Standards Institute), Европейская конференция почтовых и телекоммуникационных ведомств (CEPT — Conference of European Postal and Telecommunication Administrations), национальные институты стандартизации. В США это ANSI, в Японии — JESA (Japanese Engineering Standards Association) в Великобритании — Министерство почт и телекоммуникаций (MPT — Ministry of Posts and Telecommunications). Весьма значимы различные промышленные ассоциации и объединения, такие как Ассоциация электронной промышленности США EIA (Electronics Industries Association) и Институт инженеров по электротехнике и электронике IEEE (США). Очевидно, что эти организации сами по себе никаких технологий не разрабатывают, но они организуют и координируют всю работу, необходимую для выбора оптимального варианта из нескольких предложенных, его доработки, документального оформления, утверждения, разрешения конфликтов и т. п. Как правило, для каждого будущего стандарта создается рабочая группа, в которую входят представители всех заинтересованных сторон, которая и занимается проработкой вопроса.

Особняком в перечне стандартизирующих организаций стоит Институт инженеров по электротехнике и электронике — IEEE (Institute of Electrical and Electronics Engineers). Достаточно сказать, что членами IEEE являются ANSI и ISO. IEEE выпускает свои собственные стандарты, имеющие общемировое значение. Как правило, они затем утверждаются ISO и/или ITU, но это уже формальность.

Безусловно, IEEE — явление американское. Но сама история этой организации показывает, насколько важную роль она играла и играет в развитии телекоммуникаций в мировом масштабе, насколько ее структура отражает реалии этой столь динамичной индустрии. Корни IEEE уходят в позапрошлый век. 13 мая 1984 года по инициативе 25 наиболее авторитетных специалистов-электротехников США, среди которых были Т. Эдисон, Э. Томсон и Э. Хьюстон, был учрежден Американский институт инженеров-электротехников AIEE (American Institute of Electrical Engineers). В области телекоммуникаций эта организация занималась вопросами проводной электросвязи. Именно AIEE принадлежит заслуга в создании первых стандартов США в данном направлении. В 1907 году в Бостоне было организовано Общество инженеров по беспроводной телеграфии (Society of Wireless Telegraph Engineers — SWTE). Через два года в Нью-Йорке создали Беспроводной институт (The Wireless Institute — TWI). Однако к 1912 году обе эти организации оказались в кризисе и решили объединиться. В результа-

те был создан Институт радиоинженеров IRE (Institute of Radio Engineers). Эта организация, ведающая вопросами стандартизации в беспроводной связи, в 1963 году объединилась с AIEE. Так 1 января 1963 года родился IEEE. Отметим, что, несмотря на международный статус, IEEE тяготеет к североамериканскому континенту. Однако это не мешает ему быть законодателем мировой моды в области телекоммуникаций. В рамках темы настоящей книги особого внимания заслуживает комитет IEEE 802, занимающийся вопросами стандартизации технологий сетей передачи данных. Усилиями этого комитета оформились в виде стандартов такие технологии, как Ethernet, Token Ring, беспроводной Ethernet и др.

В России вопросами стандартизации в качестве головного национального института занимался и продолжает заниматься Госстандарт (в сотрудничестве с отраслевыми институтами — ЦНИИС, ЛОНИИС и др.). К сожалению, работа Госстандарта в области телекоммуникаций в плане участия в международной стандартизационной деятельности не отличается оперативностью. Возможно, в этом есть и положительная сторона, поскольку сегодня отечественные специалисты оперируют международными стандартами — ISO, ITU, IEEE. Это гарантирует, что не появится очередной стандарт, либо не согласующийся с общемировой практикой (как до сих пор действующий в области телефонии ГОСТ 7153-85 с уникальным требованием на сопротивление наборного ключа телефонного аппарата при замыкании шлейфа менее 50 Ом), либо с явно не лучшими техническими решениями (вспомним частотную систему сигнализации «2 из 6» с равномерным шагом между частотами 200 Гц в диапазоне 700–1700 Гц, что требует существенно разнополосных фильтров). Но по мере интеграции с мировым телекоммуникационным сообществом ситуация, надеемся, будет меняться к лучшему.

1.4 Модель взаимодействия открытых систем

Эталонная модель взаимодействия открытых систем (MBOC, OSI — open system interconnection) — это наиболее удачная попытка стандартизовать протоколы обмена информацией. Она была разработана и утверждена ISO в тесном взаимодействии с ССИТТ в 1984 году. MBOC не только стала основой для разработки сетевых стандартов, но и явилась хорошей методологической основой для изучения и сравнения сетевых технологий. Несмотря на то что были разработаны и другие модели, большинство разработчиков и поставщиков сетевых продуктов используют терминологию эталонной модели MBOC.

В соответствии с MBOC все протоколы взаимодействия систем подразделяются на семь уровней — физический, канальный (звена данных), сетевой, транспортный, сеансовый, представительский и прикладной. Рассмотрим кратко основные функции перечисленных уровней.

Нижним уровнем иерархии является физический, определяющий электрические и механические характеристики подключения к физическим каналам связи, а также процедуры передачи потока битов от одного узла к другому. Этот уровень обеспечивает сервис для канального уровня, определяющего функции управления передачей информации по каналу связи между двумя узлами. К ним,

в первую очередь, относятся упаковка передаваемой информации в кадры определенной длины, формирование проверяющих символов и проверка содержимого кадров после их передачи, передача и прием подтверждений о приеме кадров, повторная передача неподтвержденных кадров и т. д. Таким образом, основные функции управления каналом связи состоят в установлении, поддержании и разъединении каналов. При этом методы управления существенно зависят от типа канала связи (физическая линия, телефонный канал, спутниковый канал и т. д.).

Сетевой уровень обеспечивает взаимодействие между узлом и сетью. Он формирует сетевые адреса пакетов, управляет потоками, адресацией, маршрутизацией, организацией и поддержанием транспортных соединений. Единицей информации протоколов сетевого уровня является пакет, поэтому иногда этот уровень называют пакетным.

Транспортный уровень предназначен для трансляции массивов из одного порта в другой. Под портом понимается конец логического канала сети передачи данных, где фактически завершаются операции транспортировки данных и начинаются операции между вычислительными процессами. Транспортный уровень обеспечивает установление и разъединение транспортных соединений, управление информационными потоками от порта до порта, сборку и разборку пакетов, принадлежащих передаваемому в сеансе связи массиву. В связи с тем, что транспортный уровень предназначен для пересылки данных из источника адресату, протоколы данного уровня часто называют межконцевыми или сквозными. Транспортный уровень является последним в иерархии уровней, обеспечивающих транспортный сервис; он освобождает более высокие уровни от организации передачи данных.

Основное назначение сеансового уровня — организация, поддержание и окончание сеансов между прикладными процессами (организация и поддержание логической связи между распределенными работами). Сеансы устанавливаются через представительский уровень. Целью представительского уровня является преобразование данных в форму, удобную для прикладной программы. На представительском уровне происходит преобразование форматов данных и команд. Прикладной уровень представляет собой процесс обработки информации — прикладные процессы. Он обеспечивает работу прикладной программы так же, как если бы она выполнялась не через базовую сеть передачи данных, а автономно в вычислительной машине.

Несмотря на различие функций протоколов четырех нижних уровней, обеспечивающих транспортный сервис, они обладают некоторыми общими характеристиками. В первую очередь это касается способов подтверждения правильности доставки информации. Надежная передача пакетов (на сетевом уровне) или кадров (на канальном уровне) осуществляется с помощью передачи подтверждений (квитанций) об успешной доставке и использовании механизмов окна и интервала ожидания (time-out). При этом, как правило, источник может послать не более N пакетов (кадров), не дожидаясь подтверждения от адресата (N — размер окна). Получение квитанции, подтверждающей прием адресатом, разрешает передачу следующего пакета (продвигает окно). В протоколах, предусматривающих передачу очередного (кадра) только после получения квитанции на предыдущий пакет, очевидно, размер окна равен единице. Время ожидания

подтверждения об успешной доставке ограничивается некоторой величиной *time-out*. В случае непоступления квитанции за время *time-out* (это может произойти из-за искажения кадра или квитанции в канале или из-за отсутствия свободного буферного пространства в принимающем устройстве) передача пакета (кадра) повторяется. Следует отметить, что описанные выше механизмы надежной доставки информации эффективно используется также для реализации процедур управления потоком.

1.5 Методы доступа к среде в беспроводных сетях

Одна из основных проблем построения беспроводных систем — это решение задачи доступа многих пользователей к ограниченному ресурсу среды передачи. Существует несколько базовых методов доступа (их еще называют методами уплотнения или мультиплексирования), основанных на разделении между станциями таких параметров, как пространство, время, частота и код. Задача уплотнения — выделить каждому каналу связи пространство, время, частоту и/или код с минимумом взаимных помех и максимальным использованием характеристик передающей среды.

Уплотнение с пространственным разделением основано на разделении сигналов в пространстве, когда передатчик посылает сигнал, используя код s_i , время t и частоту f в области s_i . То есть каждое беспроводное устройство может вести передачу данных только в границах одной определенной территории, на которой любому другому устройству запрещено передавать свои сообщения.

К примеру, если радиостанция вещает на строго определенной частоте на закрепленной за ней территории, а какая-либо другая станция в этой же местности также начнет вещать на той же частоте, то слушатели радиопередач не смогут получить «чистый» сигнал ни от одной из этих станций. Другое дело, если радиостанции работают на одной частоте в разных городах. Искажений сигналов каждой радиостанции не будет в связи с ограниченной дальностью распространения сигналов этих станций, что исключает их наложение друг на друга.

Еще недавно данный метод считался малоэффективным — до тех пор, пока не получили промышленное развитие системы, обеспечивающие достаточно точную локализацию зон действия отдельных передатчиков. С появлением аппаратуры (и соответствующих стандартов), обеспечивающей адаптивную перестройку мощности передатчиков абонентских и базовых станций, а также систем на основе секторных антенн (или антенн с перестраиваемой диаграммой направленности) данный метод получил широкое распространение. Характерный пример — системы сотовой телефонной связи.

В схемах **уплотнения с частотным разделением** (Frequency Division Multiplexing, FDM) каждое устройство работает на строго определенной частоте, благодаря чему несколько устройств могут вести передачу данных на одной территории. Это один из наиболее известных методов, так или иначе используемый в самых современных системах беспроводной связи.

Наглядная иллюстрация схемы частотного уплотнения — функционирование в одном городе нескольких радиостанций, работающих на разных частотах. Для

надежной отстройки друг от друга их рабочие частоты должны быть разделены защитным частотным интервалом, позволяющим исключить взаимные помехи.

Эта схема, хотя и позволяет использовать множество устройств на определенной территории, сама по себе приводит к неоправданному расточительству обычно скудных частотных ресурсов, поскольку требует выделения отдельной частоты для каждого беспроводного устройства.

Более гибкой схемой является **уплотнение с временным разделением** (Time Division Multiplexing, TDM). В данной схеме распределение каналов идет по времени, т. е. каждый передатчик транслирует сигнал на одной и той же частоте f в области s , но в различные промежутки времени t_i (как правило, циклически повторяющиеся) при строгих требованиях к синхронизации процесса передачи.

Подобная схема достаточно удобна, так как временные интервалы могут динамично перераспределяться между устройствами сети. Устройствам с большим трафиком назначаются более длительные интервалы, чем устройствам с меньшим объемом трафика.

Однако метод временного уплотнения не может использоваться в чисто аналоговых сетях — даже если исходные данные аналоговые (например, речь), он требует их оцифровки и разбиения на пакеты. Скорость передачи отдельного пакета, как правило, существенно превосходит скорость передачи исходных оцифрованных данных. Характерный пример применения временного уплотнения (в проводных сетях) — это метод магистральной передачи телефонного трафика посредством каналов E1. На узловой АТС каждый аналоговый телефонный канал преобразуется в поток данных со скоростью 64 кбит/с (8 разрядов оцифровки $\times 8$ кГц частоты выборки). Фрагменты по 8 бит из 32 каналов (30 телефонных и 2 служебных) образуют цикл. Длительность каждого цикла — 125 мкс, соответственно, скорость передачи данных — $(32 \times 8 \text{ бит}) / 125 \text{ мкс} = 2048 \text{ кбит/с}$ (т. е. 2048 000 бит/с). Данный поток транслируется по магистральным каналам и восстанавливается (демультиплексируется) на приемном конце.

Основной недостаток систем с временным уплотнением — это мгновенная потеря информации при срыве синхронизации в канале, например, из-за сильных помех, случайных или преднамеренных. Однако успешный опыт эксплуатации таких знаменитых TDM-систем, как сотовые телефонные сети стандарта GSM, свидетельствует о достаточной надежности механизма временного уплотнения.

Последний тип мультиплексирования — это **уплотнение с кодовым разделением** (Code Division Multiplexing, CDM). Первоначально, из-за сложности реализации, данная схема использовалась в военных целях, но со временем прочно заняла свое место в гражданских системах. Именем основанного на CDM механизма разделения каналов (CDMA — CDM Access) даже назван стандарт сотовой телефонной связи IS-95a, а также ряд стандартов третьего поколения сотовых систем связи (cdma2000, WCDMA и др.). В данной схеме все передатчики передают сигналы на одной и той же частоте f в области s во время t , но с разными кодами c_i .

Принцип кодового уплотнения иллюстрирует ситуация, когда много людей в одной комнате разговаривают на разных языках. При этом каждый человек понимает только один определенный язык. Для каждого речь на непонятном языке будет восприниматься как ничего не значащий шум, лишенный полезной

информации. А на фоне этого шума он будет воспринимать поток информации на понятном ему языке.

В схеме CDM каждый передатчик заменяет каждый бит исходного потока данных на CDM-символ — кодовую последовательность длиной в 11, 16, 32, 64 и т. п. бит (их называют чипами). Кодовая последовательность уникальна для каждого передатчика, причем их подбирают так, чтобы корреляция двух любых CDM-кодов была минимальна (а в ряде случаев — чтобы автокорреляция CDM-кода при фазовом сдвиге была максимально возможной). Как правило, если для замены «1» в исходном потоке данных используют некий CDM-код, то для замены «0» применяют тот же код, но инвертированный.

Приемник знает CDM-код передатчика, сигналы которого должен воспринимать. Он постоянно принимает все сигналы, оцифровывает их. Затем в специальном устройстве (корреляторе) производит операцию свертки (умножения с накоплением) входного оцифрованного сигнала с известным ему CDM-кодом и его инверсией. В несколько упрощенном виде это выглядит как операция скалярного произведения вектора входного сигнала и вектора с CDM-кодом. Если сигнал на выходе коррелятора превышает некий установленный пороговый уровень, приемник считает, что принял 1 или 0. Для увеличения вероятности приема передатчик может повторять посылку каждого бита несколько раз. При этом сигналы других передатчиков с другими CDM-кодами приемник воспринимает как аддитивный шум. Более того, благодаря большой избыточности (каждый бит заменяется десятками чипов) мощность принимаемого сигнала может быть сопоставима с интегральной мощностью шума. Похожести CDM-сигналов на случайный (гауссов) шум добиваются, используя CDM-коды, порожденные генератором псевдослучайных последовательностей. Такие кодовые последовательности называют шумоподобными, соответственно, модулированные ими сигналы — шумоподобными сигналами сигналов (ШПС). Очевидно, что при передаче посредством ШПС спектр исходного сообщения расширяется во много раз. Поэтому данный метод еще называют методом расширения спектра сигнала посредством прямой последовательности (DSSS — Direct Sequence Spread Spectrum).

Наиболее сильная сторона данного уплотнения заключается в повышенной защищенности и скрытности передачи данных: не зная кода, невозможно получить сигнал, а в ряде случаев — и обнаружить его присутствие. Кроме того, кодовое пространство несравненно более значительно по сравнению с частотной схемой уплотнения, что позволяет без особых проблем присваивать каждому передатчику свой индивидуальный код. Основной же проблемой кодового уплотнения до недавнего времени являлась сложность технической реализации приемников и необходимость обеспечения точной синхронизации передатчика и приемника для гарантированного получения пакета.

Отметим, что уплотнение с кодовым разделением — метод синтетический, т. е. он базируется на частотном либо временном методе уплотнения. В наиболее «чистом» виде метод кодового уплотнения реализуется в случае DSSS. Кроме того, известны и используются методы расширения спектра посредством частотных и временных скачков (соответственно FHSS — Frequency Hopping Spread Spectrum и THSS — Time Hopping Spread Spectrum). В случае расширения спектра посредством частотных скачков в заданном частотном диапазоне F одновременно рабо-

тает несколько передатчиков, каждый в узкой полосе, во много раз меньшей F . Центральная частота каждого передатчика в ходе работы дискретно изменяется по закону, задаваемому уникальной для него кодовой последовательностью. Приемник знает эту кодовую последовательность и перестраивается по частоте приема синхронно с передатчиком. Кодовые последовательности выбирают так, чтобы минимизировать вероятность одновременной работы двух передатчиков. Тем самым обеспечивается определенная защита от прослушивания и помех. Данный метод в ряде случаев оказывается достаточно эффективным и применяется, в частности, в такой популярной сегодня технологии БСПИ, как Bluetooth.

Если метод частотных скачков представляет из себя метод частотного уплотнения с изменением частотной полосы, то метод временных скачков аналогичен временному уплотнению, только моменты начала трансляции пакетов передатчика не строго периодичны, а изменяются по псевдослучайному закону. Как правило, кодовая последовательность определяет время отклонения начала трансляции очередного пакета от заданного периода. Подобный механизм, в частности, реализован в системах связи со сверхширокой спектральной полосой компании Time Domain.

Еще одна важная производная методов кодового и частотного уплотнения — механизм мультиплексирования посредством ортогональных несущих (OFDM — Orthogonal Frequency Division Multiplexing). Его суть: весь доступный частотный диапазон разбивается на достаточно много поднесущих (от нескольких сот до тысяч). Одному каналу связи (приемнику и передатчику) назначают для передачи несколько таких несущих, выбранных из всего множества по определенному закону. Передача ведется одновременно по всем поднесущим, т. е. в каждом передатчике исходящий поток данных разбивается на N субпотоков, где N — число поднесущих, назначенных данному передатчику. Распределение поднесущих в ходе работы может динамически изменяться, что делает данный механизм не менее гибким, чем метод временного уплотнения.

До недавнего времени распространение технологии OFDM сдерживала сложность его аппаратной реализации. Однако сегодня, с развитием полупроводниковой технологии, это уже не является преградой. В результате, сегодня метод OFDM приобретает все большее распространение, в частности используется в системах связи таких популярных стандартов, как IEEE 802.16 a/g и DVB, а также является одним из основных механизмов стандарта широкополосных региональных БСПИ IEEE 802.16-2004.

Как правило, описанные схемы в беспроводных сетях используются в сочетании друг с другом. Например, для мобильных сетей GSM одновременно используются схемы уплотнения SDM, TDM и FDM, в системах стандарта IEEE 802.16 эффективно сочетаются технологии OFDM, CDM, FDM/TDM и SDM.

Рассмотренные выше механизмы — это способы разделения единого ресурса на каналы передачи. Однако эти каналы надо еще назначить конкретным устройствам. Рассмотрим несколько наиболее популярных схем распределения канальных ресурсов на базе технологии TDM (аналогичные механизмы возможны и при других методах уплотнения).

Простейший алгоритм для схемы уплотнения TDM — это **фиксированное распределение временных интервалов** между различными устройствами.

Распределением занимается базовая станция (центральное устройство), которая сообщает каждому абонентскому устройству время начала передачи. Подобная схема идеально подходит для беспроводных сетей, которые имеют фиксированную пропускную способность. Однако она не оптимальна в случае нерегулярной передачи, поскольку во время молчания устройства его пустой интервал не может быть использован другим терминалом. Поэтому число абонентских станций (либо допустимая скорость передачи) принципиально и существенно ограничено.

Противоположностью данной схемы является механизм полностью случайного доступа или **классическая схема Aloha**. В ней при передаче данных мобильным устройством не используется какой-либо алгоритм, который позволял бы избежать коллизий (одновременной работы двух передатчиков в одно время на одной частоте). Это означает, что любое устройство может передавать данные в любое время и нет никакой гарантии, что эти данные будут успешно доставлены получателю. Данная схема — один из самых первых механизмов доступа для систем беспроводной связи. Она была разработана в 70-х годах в Гавайском университете и применялась в сети ALOHANET для беспроводного соединения нескольких станций (университетских зданий, располагавшихся на разных островах Гавайского архипелага). Данная схема хорошо работает в сетях со слабой загрузкой, т. е. в сетях, имеющих малое число устройств или передающих небольшое количество информации в единицу времени. При пуассоновском распределении интенсивности генерации пакетов устройствами максимальная пропускная способность системы достигается уже при 18% загрузке.

Усовершенствованием основной схемы Aloha явился **метод множественного доступа с детектированием несущей** (Carrier Sense Multiple Access, CSMA). Детектирование несущей частоты означает лишь то, что канал прослушивается устройством. Если он занят, т. е. другое устройство передает данные, то передатчик переходит в ждущий режим до того момента, когда канал станет свободным. Этот метод позволяет значительно улучшить пропускную способность системы. Как и в методе случайного доступа, в данной схеме не требуется наличия центрального устройства, т. е. каждое устройство принимает решение о передаче самостоятельно. Поскольку фактически доступ к среде получает та станция, которая первой начала передачу, данный механизм еще называют методом конкурентного доступа.

Существует несколько версий схемы CSMA. При использовании неустойчивой схемы CSMA станции слушают канал и, если канал свободен, немедленно начинают передачу. Если канал занят, станция перед повторным определением состояния канала выжидает случайный промежуток времени, после чего опять слушает канал. Если канал свободен, то терминал передает данные. В p -настойчивых схемах CSMA узлы тоже определяют состояние канала, но данные передаются с вероятностью p . Устройство может отложить передачу до следующего временного интервала с вероятностью $1 - p$, т. е. осуществляется дополнительное разделение доступа к среде. В l -настойчивых системах CSMA все станции, которым необходимо передавать данные, одновременно получают доступ к среде, как только она освобождается.

Другой вариацией данного метода является CSMA/CA (CA — Collision Avoidance, с предотвращением конфликтов), используемая в беспроводных ЛВС

стандарта IEEE 802.11. Здесь после определения занятости канала время ожидания выбирается случайно в некотором временном промежутке. В спецификации HIPERLAN 1 используется схожая схема — беспriorитетный множественный доступ с исключением (Elimination Yield — Non-Preemptive Multiple Access, EY — NPMA).

Схема с цифровым детектированием (DSMA, Digital Sense Multiple Access) использует схожий с CSMA/CA принцип работы. Этот метод также называют множественным доступом с детектированием подавления (Inhibit Sense Multiple Access, ISMA). Различие заключается в том, что занятость канала определяется не путем прослушивания, а посредством посылки базовой станцией пакета, в котором определяется статус канала. В данной схеме базовая станция должна быть синхронизирована с передатчиками так, чтобы передатчики не передавали данные во время передачи статуса канала. Если канал занят, то станции ждут случайного промежутка времени для последующей передачи. Поскольку несколько станций могут одновременно передать данные, центральная станция посылает пакет с подтверждением о получении пакета данных.

В современных БСПИ, как правило, используют сочетание механизмов централизованного назначения временных интервалов и методов конкурентного доступа. По сути, работа этих систем происходит в два этапа. Первый этап — резервирование ресурсов (временных интервалов) для будущей передачи. На этом этапе все станции заявляют (пытаются заявить) о своих потребностях в канальных ресурсах. На втором этапе происходит непосредственная передача данных в отведенном временном интервале. В этих схемах используется центральный терминал, с помощью которого производится синхронизация передач и осуществляется резервирование. Как правило, механизмы резервирования приводят к увеличению времени задержки получения пакетов при слабой загрузке системы, но при этом обеспечивают ей более высокую пропускную способность.

Примером подобного механизма является **схема множественного доступа с распределением по запросу (Demand Assigned Multiple Access, DAMA)**, называемая также схемой Aloha с резервированием. Она, в частности, применяется в спутниковых системах связи. В течение определенного временного интервала, разбитого на мини-интервалы, все станции пытаются зарезервировать для себя будущие временные интервалы для передачи данных. Поскольку на стадии резервирования происходят конфликты, некоторые станции не резервируют для себя временные интервалы. Если станции удалось зарезервировать временной интервал, то ни одна другая станция не сможет в это время осуществлять передачу. Таким образом, базовая станция собирает все успешные запросы (остальные игнорируются) и посылает назад список с указанием прав доступа к последующим временным интервалам. Этому списку подчиняются все станции. Схема DAMA относится к схемам с явным резервированием, когда каждый интервал для передачи резервируется явно.

Схема TDMA с резервированием отличается от предыдущей схемы тем, что этап резервирования происходит не на основании конкурентного доступа, а по обычной фиксированной схеме TDMA. Каждому устройству назначается временной мини-интервал, в течение которого оно сообщает, будет ли передавать данные. Поэтому в начале каждого цикла передачи базовая станция передает

пакет, разбитый на N интервалов, в каждом из которых указано, зарезервирован канал или нет. Затем следуют $N \cdot k$ интервалов для данных. Данный метод гарантирует каждой зарезервировавшей канал станции определенную пропускную способность. Остальные станции могут пересылать данные в течение интервалов, которые никто не зарезервировал, но уже на принципах конкурентного доступа и без гарантии доставки пакетов.

Схема с резервированием пакетов (PRMA, Packet Reservation Multiple Access) является примером со скрытым резервированием, поскольку интервалы резервируются неявно. Центральное устройство в начале каждого цикла рассылает список с распределением временных интервалов. Само же резервирование происходит по другой схеме. Представим, что какому-либо устройству необходимо передать данные, но при этом он не зарезервировал временной интервал. Это устройство регулярно получает список с зарезервированными интервалами. К примеру, в полученном списке указано, что третий, пятый и восьмой интервалы не зарезервированы, т. е. свободны. Устройство случайным образом принимает решение о том, в каком интервале можно попытаться передавать данные. Например, устройство передает сообщение в пятый интервал. Если передача прошла успешно, устройство получает об этом подтверждение. Базовая станция резервирует этот канал для нового устройства и включает его в свой список. Если запрос не дошел до базовой станции, устройство должно попробовать вновь послать данные в один из свободных интервалов.

ГЛАВА 2

Технологии и архитектура беспроводных сетей

2.1 Персональные беспроводные сети (технологии Bluetooth, Home RF, IEEE 802.15.3(4))

Персональные беспроводные сети передачи данных стали появляться сравнительно недавно — в середине 90-х годов. Однако лишь к концу 90-х годов развитие микроэлектроники позволило производить для таких устройств интегральную элементную базу. Открывшиеся перспективы привели к тому, что практически одновременно появилось сразу несколько разработок персональных БСПИ, основные из которых — это спецификации HomeRF, стандарты семейства IEEE 802.15 (Bluetooth и IEEE 802.15.3/4), а также технологии сверхширокополосной связи (UWB).

2.1.1 Стандарты Bluetooth и HomeRF

HomeRF — это название созданной в марте 1998 года группы производителей компьютерного и бытового оборудования (Home Radio Frequency Working Group; в первый год существования в нее вошло свыше 90 фирм, включая Intel, Compaq, Ericsson, Hewlett-Packard и Microsoft). Она организовывалась для разработки открытого протокола распределенного беспроводного доступа SWAP (Shared Wireless Access Protocol), который должен был лечь в основу радиосети HomeRF. Впоследствии вместо понятия SWAP в названии спецификации стали использовать HomeRF. Первая версия спецификации HomeRF появилась 17 декабря 1998 года, последняя (HomeRF 2.01) — 1 июля 2002 года. На этом цели рабочей группы были выполнены, и в январе 2003 года она была расформирована.

Как следует из названия, стандарт HomeRF задумывался как беспроводная домашняя СПИ, аналог известной к тому времени проводной Ethernet-подобной сети HomePNA (Home Phone Line Networking Alliance), в качестве носителя ис-

пользовавшей уже проложенную телефонную линию. Для нормальной работы сети HomeRF необходим host-компьютер (или устройство, выполняющее его функции).

Идеология Bluetooth иная — это универсальный радиointерфейс, связывающий друг с другом самые разные устройства и не требующий дорогой аппаратной поддержки. Однако устройства Bluetooth сегодня реально используют в основном для замены провода радиointерфейсом (например, в качестве беспроводной гарнитуры для сотовых телефонов), несмотря на широчайший спектр заложенных в нем возможностей. Рынок такого рода приложений пока во много раз превосходит рынок действительно сетевых устройств. Возможно, именно поэтому прекрасно проработанный для применения именно для сетевых задач стандарт HomeRF пока не нашел массового применения. С одной стороны его вытесняют простейшие Bluetooth-устройства, с другой — системы стандарта IEEE 802.11, которые за последние пять лет существенно подешевели, лишив тем самым HomeRF основного перед ними преимущества — низкой стоимости. Поэтому более детально рассмотрим именно стандарт Bluetooth.

Основные характеристики технологий Bluetooth и HomeRF приведены в табл. 2.1.

Таблица 2.1. Сравнительные характеристики технологий Bluetooth и Home RF

Показатель	HomeRF	Bluetooth
Вид модуляции	Шумоподобный сигнал, метод частотных скачков	Шумоподобный сигнал, метод частотных скачков
Число скачков в секунду	50	1600
Мощность передатчика, мВт	100	100
Скорость обмена данными, Мбит/с	1 или 2*	1
Способ модуляции	Двух- или четырехуровневая ЧМ	Двухуровневая ЧМ
Количество устройств в сети	До 127	Не ограничено
Защита информации	Blowfish data security	40- и 64-битное шифрование
Радиус действия, м	50	10–100

*Возможно, до 10.

2.1.2 Архитектура и логическая структура сети Bluetooth

Своим появлением спецификация Bluetooth обязана компаниям Ericsson, IBM, Intel, Toshiba и Nokia, которые в феврале 1998 года для разработки стандарта персональной БСПИ организовали специальную рабочую группу SIG (Special Interest Group). Название новой технологии связано с именем короля Харольда I Блаатанда (в поздней транскрипции — Bluetooth, Синезуб), объединившего Данию с Южной Швецией и Южной Норвегией в единое Датское Королевство.

Уже в 2000 году в Bluetooth SIG входили 1883 фирмы (на порядок больше, чем в группе HomeRF). Новую технологию поддерживали производители элементной базы, программного обеспечения, портативных компьютеров, сотовых телефонов, звуковоспроизводящей аппаратуры и др. Видимо, имя короля-объединителя принесло свои плоды. Сегодня стандарт Bluetooth признан всем мировым сообществом. Между Bluetooth SIG и IEEE было достигнуто соглашение, в соответствии с которым спецификация Bluetooth вошла в стандарт IEEE 802.15.1 (опубликован 14 июня 2002 года) «Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Wireless Personal Area Networks (WPANs)» (Спецификации контроля доступа к беспроводному каналу и физического уровня беспроводных персональных сетей).

Спецификация Bluetooth описывает пакетный способ передачи информации с временным мультиплексированием. Радиообмен происходит в полосе частот 2400–2483,5 МГц (в США и ряде других стран — безлицензионный диапазон). В радиотракте применен метод расширения спектра посредством частотных скачков и двухуровневая частотная модуляция с фильтром Гаусса (binary Gaussian Frequency Shift Keying).

Метод частотных скачков подразумевает, что вся отведенная для передачи полоса частот подразделяется на определенное количество подканалов шириной 1 МГц каждый. Канал представляет собой псевдослучайную последовательность скачков по 79 или 23 радиочастотным подканалам (табл. 2.2). Каждый канал делится на временные сегменты продолжительностью 625 мкс, причем каждому сегменту соответствует определенная несущая (подканал). Передатчик «перескакивает» с несущей на несущую синхронно с приемником в последовательности, определяемой номером канала. За секунду может происходить до 1600 частотных скачков. Такой метод обеспечивает конфиденциальность и некоторую помехозащищенность передач. Последняя обусловлена тем, что если переданный по какому-либо подканалу пакет не был принят, то приемник сообщает об этом, и передача пакета повторяется на одном из следующих подканалов, уже на другой частоте.

Таблица 2.2. Разделение полосы частот на подканалы в стандарте Bluetooth

Страна	Диапазон, МГц	Несущая частота подканалов, МГц	Допустимые номера подканалов, k
Европа* и США	2400,8–2483,5	$2402 + k$	0...79
Япония	2471,8–2497,8	$2473 + k$	0...23
Испания	2445,8–2475,8	$2449 + k$	0...22
Франция	2446,5–2483,5	$2454 + k$	0...22

*Кроме Испании и Франции.

Протокол Bluetooth поддерживает соединения типа точка-точка и точка-многоточка. Два или более использующих один и тот же канал устройства образуют пикосеть (piconet). Одно из устройств работает как основное (master), а остальные — как подчиненные (slaves). В одной пикосети может быть до семи активных

подчиненных устройств, при этом остальные подчиненные устройства находятся в состоянии «парковки», оставаясь синхронизированными с основным устройством. Взаимодействующие пикосети образуют «распределенную сеть» (scatternet).

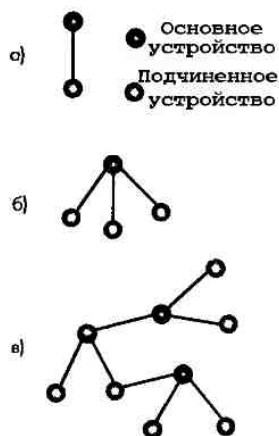


Рис. 2.1. Пикосеть с одним подчиненным устройством (а), несколькими (б) и распределенная сеть (в)

В каждой пикосети действует только одно основное устройство, однако подчиненные устройства могут входить в различные пикосети. Кроме того, основное устройство одной пикосети может являться подчиненным в другой (рис. 2.1). Пикосети не синхронизированы друг с другом по времени и частоте — каждая из них использует свою последовательность частотных скачков. В одной же пикосети все устройства синхронизированы по времени и частотам. Псевдослучайная последовательность скачков уникальна для каждой пикосети и определяется адресом ее основного устройства. Длина цикла псевдослучайной последовательности — 227 элементов.

В стандарте Bluetooth предусмотрена дуплексная передача на основе разделения времени (time division duplexing — TDD). Основное устройство передает пакеты $f(k)$ в нечетные временные сегменты, а подчиненное устройство — в четные (рис. 2.2). Пакеты в зависимости от длины могут занимать до пяти временных сегментов. При этом частота канала не меняется до окончания передачи пакета (рис. 2.3).

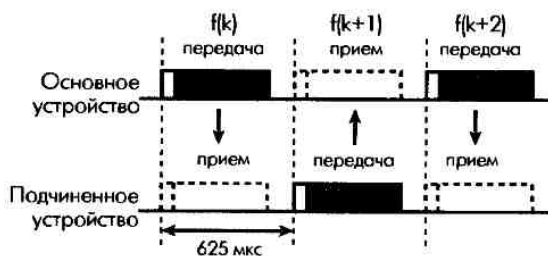


Рис. 2.2. Временные диаграммы работы канала

кбит/с в прямом и 57,6 кбит/с в обратном направлениях или до 433,9 кбит/с в каждом направлении в симметричном режиме.

Синхронное соединение (SCO) возможно только в режиме точка-точка. Такой вид связи применяется для передачи информации, чувствительной к задержкам, например голоса. Основное устройство поддерживает до трех синхронных соединений, вспомогательное — до трех синхронных соединений с одним основным устройством или до двух — с разными основными устройствами.

При синхронном соединении основное устройство резервирует временные сегменты, следующие через так называемые SCO-интервалы. Даже если пакет принят с ошибкой, повторно при синхронном соединении он не передается.

Протокол Bluetooth может поддерживать асинхронный канал данных, до трех синхронных (с постоянной скоростью) голосовых каналов или канал с одновременной асинхронной передачей данных и синхронной передачей голоса. Скорость каждого голосового канала — 64 кбит/с в каждом направлении, асинхронного в асимметричном режиме — до 723,2

При асинхронной связи (ACL) используются временные сегменты, не зарезервированные для синхронного соединения. Асинхронное соединение возможно между основным и всеми активными подчиненными устройствами в пикосети. Основное и подчиненное устройства могут поддерживать только одно асинхронное соединение. Поскольку в пикосети может быть

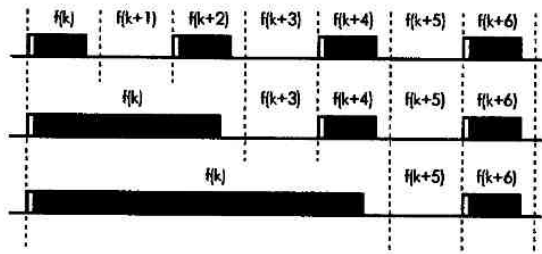


Рис. 2.3. Передача пакетов различной длины

несколько подчиненных устройств, конкретное подчиненное устройство отправляет пакет основному, только если в предыдущем временном интервале на его адрес пришел пакет от основного устройства. Если в адресном поле ACL-пакета адрес не указан, пакет считается «широковещательным» — его могут читать все устройства. Асинхронное соединение позволяет повторно передавать пакеты, принятые с ошибками (механизм ARQ — automatic repeat request).

Стандартный пакет Bluetooth содержит код доступа длиной 72 бита, 54-битный заголовок и информационное поле длиной не более 2745 бит (рис. 2.4). Однако пакеты могут быть различных типов. Так, пакет может состоять только из кода доступа (в этом случае его длина равна 68 битам) или кода доступа и заголовка.

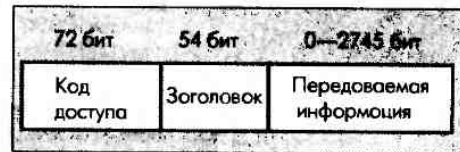


Рис. 2.4. Структура пакета

Код доступа идентифицирует пакеты, принадлежащие одной пикосети, а также используется для синхронизации и реализации процедуры запросов. Он включает преамбулу (4 бита), слово синхронизации (64 бита) и трейлер — 4 бита контрольной суммы (рис. 2.5).

Заголовок содержит информацию для управления связью и состоит из шести полей (рис. 2.6):

AM_ADDR — 3-битный адрес активного элемента (active member address);

TYPE — 4-битный код типа данных;

FLOW — 1 бит управления потоком данных, показывающий готовность устройства к приему;

ARQN — 1 бит подтверждения правильного приема;

SEQN — 1 бит, служащий для определения последовательности пакетов;

HEC — 8-битная контрольная сумма.

Информационное поле в зависимости от типа пакетов может содержать либо поля голоса, либо поля данных, либо оба типа полей одновременно.

Поскольку глобальная цель Bluetooth — всеобщее объединение, эта спецификация описывает протоколы работы практически на всех уровнях модели взаимо-

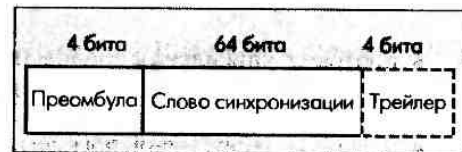


Рис. 2.5. Структура кода доступа

действия открытых систем (МВОС/ОСИ). Стек протоколов Bluetooth позволяет устройствам различных стандартов взаимодействовать друг с другом посредством данной технологии.

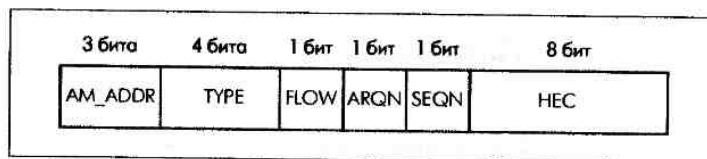


Рис. 2.6. Структура заголовка

До сих пор мы рассматривали только физический уровень и уровень звена данных, поскольку взаимодействие на этих уровнях поддерживается аппаратно. Протоколы верхних уровней, как правило, реализуются программно. Основные протоколы взаимодействия, входящие в Bluetooth (рис. 2.7):

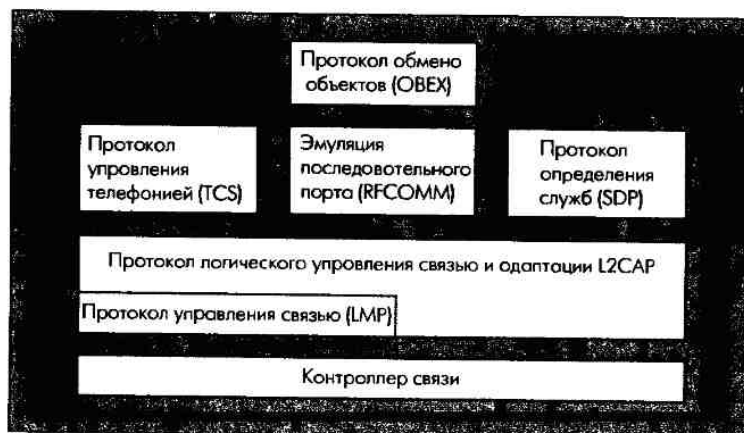


Рис. 2.7. Стек протоколов Bluetooth

- протокол управления соединением (Link manager protocol). Используется для установления связи, управления и защиты информации;
- протокол управления логическим соединением и адаптации (Logical link control and adaptation protocol L2CAP). Обеспечивает мультиплексирование, сегментацию и переконфигурацию пакетов;
- протокол определения служб (SDP). Позволяет идентифицировать тип и характеристики взаимодействующего устройства;
- протокол RFCOMM. Основан на стандарте ETSI TS 07.10, поддерживает интерфейс RS-232, обеспечивая эмуляцию последовательного порта;
- протокол управления телефонией (TCS). Служит для организации соединения между устройствами для передачи голоса и данных;

- протокол обмена объектов OBEX. Является основой для работы различных пользовательских приложений через канал Bluetooth (например, электронной почты). Он же обеспечивает совместное использование Bluetooth и других коммуникационных интерфейсов, например IrDA (Infrared Data Association).

2.1.3 Технические средства сети Bluetooth

Одно из необходимых условий успеха такой технологии, как Bluetooth, — недорогая программно-аппаратная реализация. В структуру устройств Bluetooth (рис. 2.8) входят радиомодуль-трансивер, контроллер связи (baseband-процессор) и управляющее связью устройство, собственно реализующее протоколы Bluetooth верхних уровней, а также интерфейс с терминальным устройством. Причем если трансивер и контроллер связи (в первых чипсетах для Bluetooth) — это специализированные микросхемы (интегральные или гибридные), то устройство управления связью реализуют на стандартных микроконтроллерах, сигнальных процессорах либо его функции поддерживают центральные процессоры мощных терминальных устройств (например, ноутбуков). Кроме того, в устройствах Bluetooth применяют интегральные схемы (ИС), используемые в других приложениях, поскольку диапазон 2 МГц освоен достаточно хорошо, а заложенные в Bluetooth технические решения сами по себе особой новизны не содержат (схема модуляции — широко распространена, технология расширения спектра методом частотных скачков хорошо отработана, мощность мала).

Рис. 2.8. Структура устройства Bluetooth



Поэтому неудивительно, что первые чипсеты для Bluetooth включали ИС, хорошо знакомые по другим приложениям. Так, одной из первых свое решение для Bluetooth представила фирма Philips Semiconductors, предложив комплект ИС, включающий четыре микросхемы: синтезатор UMA1022, усилитель мощности SA2410, схему модуляции SA639 и трансивер SA2420. Две из них — UMA1022 и SA639 — используются в DECT-устройствах, да и остальные достаточно универсальны. Однако вскоре Philips предложила специализированный чипсет, поддерживающий спецификацию Bluetooth 1.0. Он включает однокристалльный трансивер UAA3558, основанный на оригинальной технологии «низкой промежуточной частоты» (изначально создавался для DECT), и baseband-процессор серии VWS2600x. Процессор VWS26002 может работать и с трансивером компании Ericsson PBA 313. Он содержит ядро 32-разрядного RISC-микроконтроллера ARM7 TDMI, кодек голоса, поддерживает интерфейсы UART, USB, PCM и I²C. Изготавливается процессор по 0,25-мкм КМОП-технологии,

поэтому он характеризуется малой потребляемой мощностью и низким напряжением: 1,8–2,5 В для цифрового ядра и 2,5–3,3 В для портов ввода/вывода. Одна из последних разработок компании Philips – полностью завершённый Bluetooth-радиомодуль DGB202, построенный на базе размещённых в одном корпусе коммуникационного Bluetooth-процессора PCF87852 и трансивера UAA3559. Вскоре должен стать доступным и перспективный baseband-процессор компании Philips – PCD 87550 (рис. 2.9, 2.10).

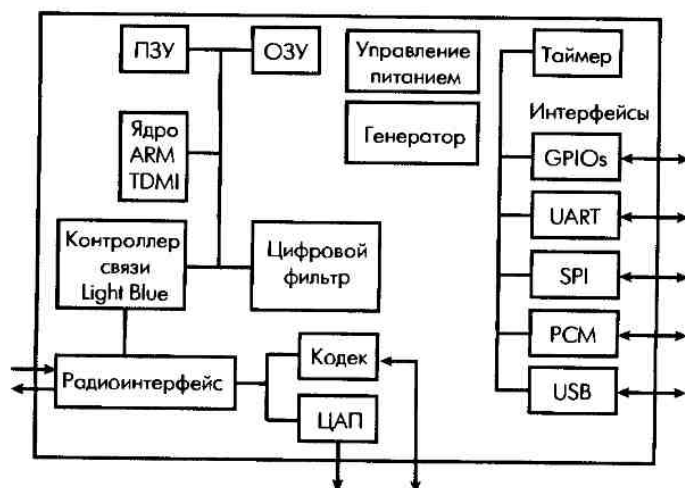


Рис. 2.9. Структурная схема Bluetooth-контроллера PCD 87550

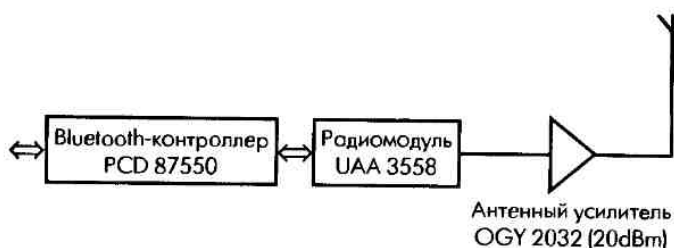


Рис. 2.10. Структура устройства Bluetooth с PCD 87550

Вообще следует отметить, что ядро ARM TDMI применяется в baseband-процессорах многих фирм. Так, оно интегрировано в Bluetooth baseband-процессор фирмы Ericsson. На основе этого ядра построен и однокристалльный контроллер Bluetooth компании Atmel AT76C551. Его структура во многом аналогична приведенной на рис. 2.9. Atmel предлагает чипсет, включающий данный контроллер и однокристалльный трансивер T2901 компании Temic Semiconductors (которая теперь вошла в состав Atmel) (рис. 2.11). Трансивер T2901 обеспечивает радиус действия до 10 м. Если его надо увеличить до 100 м, Atmel предлагает SiGe-ИС T7024, включающую малошумящий предусилитель и усилитель мощности 23 дБм (рис. 2.12).

В отличие от упомянутых выше монолитных ИС, трансивер компании Ericsson PBA31301 хотя и основан на специализированной БиКМОП ИС, но является

гибридным модулем, собранным на многослойной керамической подложке (шесть слоев металлизации). В исполнении 100 Вт размеры модуля — $10,2 \times 16 \times 1,6$ мм.

Компания Lucent Technologies также производит чипсет для Bluetooth, содержащий однокристалльный передатчик W7020 с низкой потребляемой мощностью (напряжение питания — до 2,7 В), и baseband-контроллер W7400. Свой комплект ИС Odyssey выпустила и фирма Silicon Wave. В него входят ИС радиомодема SiW1501 и контроллер связи SiW1601.

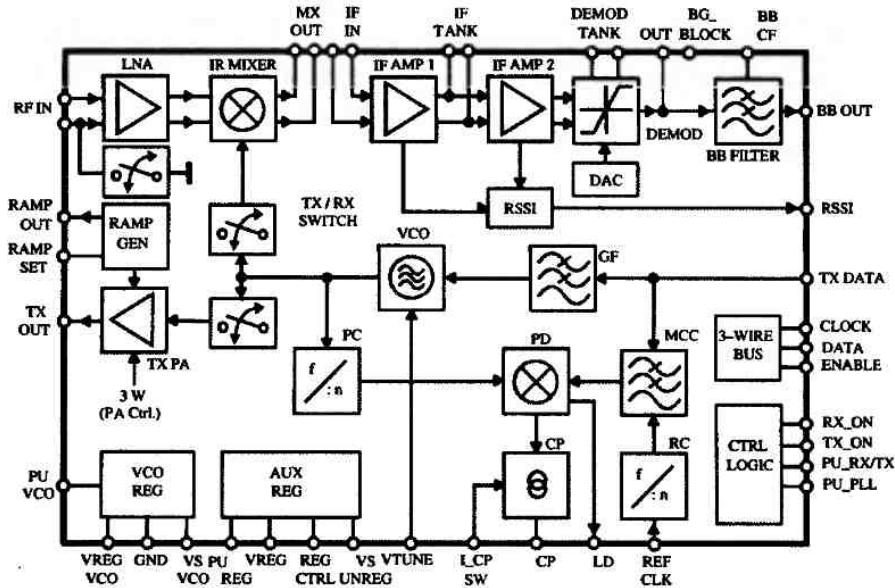


Рис. 2.11. Структура трансивера T2901

Не осталась в стороне и National Semiconductor. Ее чипсет состоит из трансивера со встроенным ФАПЧ LMX3162 и контроллера связи LMX5001. Как и в случае чипсета Odyssey, при реализации Bluetooth-устройств на базе этого комплекта схем необходим процессор, выполняющий функции управления связи. Им может быть центральный процессор компьютера или, например, сигнальный процессор ADSP-218x (Analog Devices) с соответствующим программным обеспечением.

Все говорит о том, что у беспроводной технологии Bluetooth — большое будущее. Ее интеграция с Интернетом может стать качественно новым этапом в развитии всемирной сетевой инфраструктуры. Причина этого — в совокупности достоинств новой технологии, главные из которых:

- небольшой радиус действия, что означает малую мощность передатчика и низкую потребляемую мощность;
- высокая устойчивость к интер- модуляционным помехам и отсутствию влияния устройств Bluetooth на обычную бытовую электронику;
- низкая стоимость.

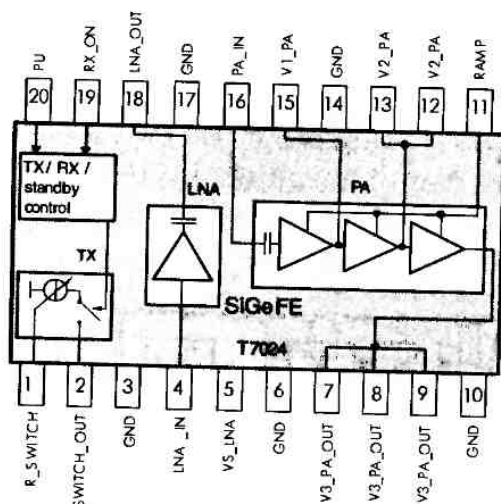


Рис. 2.12. Структура ИС T7024

Несмотря на всю свою привлекательность и универсальность, стандарт Bluetooth как действительно сетевой стандарт используется сегодня относительно редко. Возможно, причина кроется в том, что все же это еще очень новая технология, а также в том, что Bluetooth слишком универсален. В одних приложениях ему недостает скорости обмена (например, при передаче видеосигнала), в других — требуются более простые и дешевые устройства.

Чтобы разрешить эти противоречия и окончательно стандартизовать принципы организации персональных СПИ, рабочая группа IEEE 802.15, не удовлетвова-

вшись стандартом IEEE 802.15.1, создала еще две исследовательские группы (Tg3 и Tg4). В результате в конце сентября 2003 года были опубликованы два новых стандарта — IEEE 802.15.3 «Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPANs)» и IEEE 802.15.4 «Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs)» — стандарты для высокоскоростной и низкоскоростной персональных БСПИ, соответственно.

2.1.4 Высокоскоростные персональные сети стандарта IEEE 802.15.3(3a)

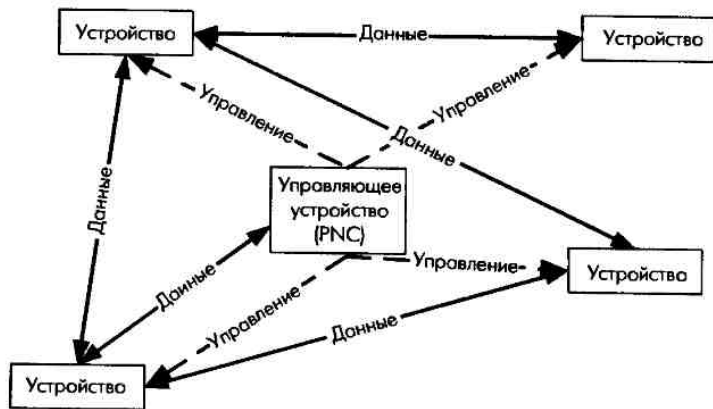
Спецификация IEEE 802.15.3

Стандарт IEEE 802.15.3 описывает работу малой БСПИ — пикосети (piconet). Пикосеть в стандарте IEEE 802.15.3 — это так называемая ad hoc — система, в которой несколько независимых устройств могут непосредственно взаимодействовать друг с другом. Радиус зоны действия одной пикосети, как правило, не превышает 10 м. Основные требования к ней — высокая скорость передачи данных, простая инфраструктура, легкость установления соединения и вхождения в сеть, наличие средств защиты данных и предоставление для определенных типов данных соединения с гарантированными параметрами передачи (гарантия качества обслуживания, QoS).

Пикосеть (рис. 2.13) может объединять несколько устройств, одно из которых выполняет функции управления (координатор пикосети — piconet coordinator, PNC). Стандарт также предусматривает возможность формирования так назы-

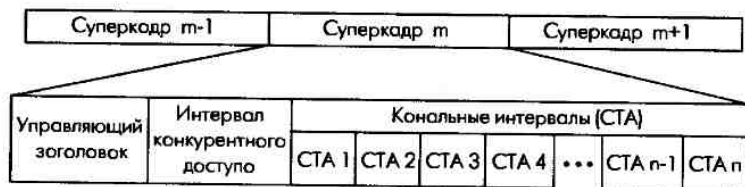
ваемых дочерних пикосетей и описывает взаимодействие между независимыми соседними пикосетями.

Рис. 2.13. Структура пикосети IEEE 802.15.3



В пикосети возможен обмен как асинхронными, так и изохронными (поточковыми) данными. К последним относятся, например, звук и видео. Весь информационный обмен в пикосети основан на последовательности суперкадров (superframe — терминология стандартов IEEE 802.15). Каждый суперкадр (рис. 2.14) включает управляющий сегмент (beacon), интервал конкурентного доступа (contention access period — CAP) и набор временных интервалов (каналов), назначенных определенным устройствам. PCN определяет границы всех интервалов и распределяет каналы между устройствами.

Рис. 2.14. Структура суперкадров физического уровня сети IEEE 802.15.3



Во время CAP доступ к каналу предоставляется на основе механизма контроля несущей с предотвращением коллизий — CSMA/CA (как и в стандарте IEEE 802.11), т.е. кто первый успел занять канал, тот и работает. В этот период передаются команды или асинхронные данные.

Канальные интервалы (СТА) координатор пикосети назначает каждому устройству или группе устройств по предварительному запросу с их стороны. В управляющем сегменте задается момент начала и длительность каждого СТА. Назначение канального интервала для какого-либо устройства означает, что никто другой в этот момент не может работать на передачу. СТА могут динамически распределяться в суперкадре (для асинхронных и изохронных данных) или быть фиксированными (только для изохронных данных).

Спецификация физического канала в документе IEEE 802.15.3 приведена только для диапазона 2400–2483,5 МГц. Она предусматривает пять допустимых скоростей передачи (табл. 2.3). Скорость 22 Мбит/с является базовой, ее обязаны поддерживать все устройства IEEE 802.15.3. При работе на этой скорости данные не кодируются. В остальных случаях данные перед формированием модуляционных символов кодируются посредством сверточного кодера с трехразрядным сдвиговым регистром (так называемая модуляция посредством решетчатого кода с восемью состояниями). При этом в кодере к исходному набору из 1/3/4/5 бит (при QPSK/16-QAM/32-QAM/64-QAM) добавляется кодовый бит с выхода трехразрядного сдвигового регистра.

Таблица 2.3. Модуляция и скорости передачи данных в сетях IEEE 802.15.3 в диапазоне 2,4 ГГц

Тип модуляции	Скорость передачи данных, Мбит/с
QPSK	11
DQPSK	22
16-QAM	33
32-QAM	44
64-QAM	55

Стандарт IEEE 802.15.3 требует, чтобы устройства могли работать в любом из пяти возможных частотных каналов (табл. 2.4). Причем предусматривается два канальных плана — режим высокой плотности (четыре канала в допустимом диапазоне) и режим совместимости с сетью стандарта IEEE 802.11b (три разрешенных канала). Это означает, что каждое устройство перед началом работы сканирует диапазон, находит свободные каналы, определяет наличие работающей сети 802.11b.

Таблица 2.4. Распределение каналов в сетях IEEE 802.15.3

Номер канала	Центральная частота, МГц	Режим высокой плотности	Режим совместимости с IEEE 802.11b
1	2412	✓	✓
2	2428	✓	
3	2437		✓
4	2445	✓	
5	2462	✓	✓

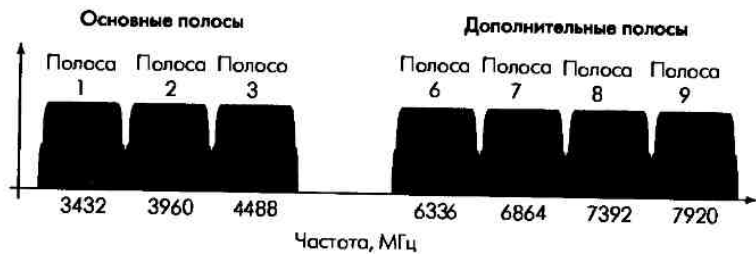
Сверхбыстродействующие персональные сети (IEEE 802.15.3a)

Спецификацию IEEE 802.15.3 не успели утвердить (а произошло это 12 июня 2003 года), как весь телекоммуникационный мир стал ждать появления нового стандарта — IEEE 802.15.3a. Речь идет о разработке принципов построения пикосети со скоростью обмена 110–480 Мбит/с и выше — до 1320 Мбит/с. Достичь столь высоких скоростей можно только увеличивая спектральную ширину

канала, переходя в область так называемой сверхширокополосной связи (СШП, UWB). В США это стало возможным после 14 февраля 2002 года, когда федеральная комиссия связи (FCC) США разрешила применение сверхширокополосных устройств внутри помещений в диапазоне 3100–10600 МГц при максимальной плотности мощности излучения $7,41 \cdot 10^{-14}$ Вт/Гц ($-41,3$ дБм/МГц).

В результате в 2002 году образовалась исследовательская группа Tg3a, в которую вошли представители практически всех крупнейших полупроводниковых и телекоммуникационных фирм. Вскоре появились два конкурирующих предложения по технологии СШП-передачи — на основе ортогональных кодов (так называемый мультиполосный множественный доступ посредством ортогональных несущих, MB-OFDM) и путем расширения спектра сигнала методом прямой последовательности (DS-UWB). Первое предложение поддерживало большинство фирм во главе с гигантами Texas Instruments и Intel (в марте 2003 года было создано даже специальное объединение MBOA — Multiband OFDM Alliance). Лагерь сторонников второго возглавили компании Motorola и XtremeSpectrum. По принятым в комитете IEEE 802 правилам, для того чтобы утвердить стандарт, за предложенный вариант должны проголосовать не менее 75% членов рабочей группы. Однако, несмотря на численный перевес сторонников MB-OFDM, а в MBOA входит 170 компаний, среди которых — 9 из 10 крупнейших полупроводниковых компаний (кроме TSMC), на прошедших голосованиях им не удалось набрать необходимые 75% голосов от общего числа компаний, работающих над стандартом IEEE 802.15.3a. Причин тут несколько. Возможно, одна из основных кроется в технологии MB-OFDM.

Рис. 2.15. Предлагаемое распределение каналов в стандарте IEEE 802.3a в США



Суть ее состоит в том, что весь разрешенный диапазон делится на полосы шириной 528 МГц. В стандартном режиме предусмотрено три полосы, в расширенном — семь (рис. 2.15). Каждая полоса, в свою очередь, делится на 128 поднесущих частот с шагом 4,125 МГц. Из них используется 122: 100 для модуляции данных, 12 поднесущих — пилотные и еще 10 — защитные. Каждая поднесущая модулируется посредством QPSK. Один OFDM-символ содержит 100 или 200 кодированных бит (100 в случае, когда одинаково модулируются две поднесущие, симметричные относительно центральной). Период следования символов — 312,5 нс. До сих пор все соответствует обычной OFDM. Мультиполосность означает, что последующий символ может передаваться в иной частотной полосе по жестко определенной схеме для каждого логического канала.

Последовательность перехода с одной полосы частот на другую называют частотно-временным кодом. Пока предусмотрено четыре таких кода (канала)

(табл. 2.5). Кроме перехода с частоты на частоту, предусмотрен режим, когда один символ может передаваться несколько раз (два или четыре). Например, код $1 \rightarrow 2 \rightarrow 3 \rightarrow 1 \rightarrow 2 \rightarrow 3$ означает, что первый OFDM-символ передается в полосах 1 и 2, второй OFDM-символ — в полосах 3 и 1, третий — в полосах 2 и 3.

Таблица 2.5. Частотно-временные коды

Номер логического канала	Режим трех полос	Режим семи полос
1	$1 \rightarrow 2 \rightarrow 3 \rightarrow 1 \rightarrow 2 \rightarrow 3$	$1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 6 \rightarrow 7$
2	$1 \rightarrow 3 \rightarrow 2 \rightarrow 1 \rightarrow 3 \rightarrow 2$	$1 \rightarrow 7 \rightarrow 6 \rightarrow 5 \rightarrow 4 \rightarrow 3 \rightarrow 2$
3	$1 \rightarrow 1 \rightarrow 2 \rightarrow 2 \rightarrow 3 \rightarrow 3$	$1 \rightarrow 4 \rightarrow 7 \rightarrow 3 \rightarrow 6 \rightarrow 2 \rightarrow 5$
4	$1 \rightarrow 1 \rightarrow 3 \rightarrow 3 \rightarrow 2 \rightarrow 2$	$1 \rightarrow 3 \rightarrow 5 \rightarrow 7 \rightarrow 2 \rightarrow 4 \rightarrow 6$

Обмен информацией происходит кадрами (пакетами). Кадр состоит из преамбулы (набора синхронизирующих последовательностей), заголовка (управляющая информация) и поля данных. Преамбула и заголовок всегда транслируются с наименьшей из возможных скоростей — 55 Мбит/с. Четыре логических канала подразумевают, что в непосредственной близости могут работать по крайней мере четыре пикосети. Для этого каждому логическому каналу соответствует уникальный вид синхропоследовательности в преамбуле.

Таким образом, предлагается комбинация OFDM и известного механизма частотных скачков (FH). В результате, в зависимости от скорости кодирования и числа повторов символов и формируется спектр скоростей от 55 до 480 Мбит/с. Изменяя вид модуляции, можно достичь и больших скоростей. Так, применение 16-QAM при той же схеме кодирования даст уже $480 \times 2 = 960$ кбит/с. Другой путь заключается в использовании для передачи одновременно трех диапазонов — тогда при QPSK и скорости сверточного кодирования $3/4$ достигается скорость $480 \times 3 = 1440$ Мбит/с.

Проблема в том, что метод частотных скачков — не самый эффективный с точки зрения использования спектрального диапазона. Оппоненты MB-OFDM указывают, что сторонники данной технологии в своих измерениях средней мощности излучения передатчика усредняют ее по временному интервалу порядка 1 мс. Это время соответствует длительности трех символов. Реально же при использовании механизма повторов (и при измерениях) символ в одном субканале за это время передается только один раз. В результате плотность мощности излучения при работе передатчика может превосходить допустимые (Федеральной комиссией связи США) — 41,3 дБм/МГц. А это уже серьезная проблема, поскольку речь идет о сигнале с полосой выше 500 МГц.

Сторонники технологии DS-UWB предлагают для расширения спектра классический метод прямой последовательности. При этом каждый бит заменяется специальной кодовой последовательностью длиной до 24 бит. Предусмотрено два

вида модуляции — двоичная фазовая BPSK (один бит на символ) и так называемая 4ВОК-модуляция (модуляция на основе четырех ортогональных двоичных кодов. Известный вариант ВОК-метод ССК, основной вид модуляции в стандарте IEEE 802.11b). 4ВОК — фактически вариант квадратурной модуляции, один 4ВОК-символ содержит 2 бита.

Весь диапазон вещания разбит на две зоны: 3,1–4,85 ГГц (нижний диапазон) и от 6,9 до 9,7 ГГц (верхний диапазон). В каждом диапазоне предусмотрено по шесть каналов пикосети (с шагом 39 МГц в нижнем диапазоне начиная с 3900 МГц и с шагом 78 МГц — в верхнем начиная с 7800 ГГц). Только четыре канала нижнего диапазона с центральными частотами 3939, 3978, 4017 и 4056 МГц считаются обязательными для поддержки каждым устройством, остальные каналы — дополнительные. Частота следования модуляционных символов в каждом канале равна 1/3 его центральной частоты. В зависимости от скорости предварительного кодирования, вида модуляции и длины кодовой последовательности скорость передачи данных может составить 28, 55, 110, 220, 500, 660, 1000 и 1320 Мбит/с.

Отметим, что споры сторонников двух подходов к реализации СШП-пикосетей длятся с 2003 года. Производители СВИС готовы начать выпуск необходимых компонентов (а некоторые уже начали), и ситуация должна вскоре разрешиться. Тем более что уже появляются предложения, как объединить две эти технологии и без особых затрат производить двухмодовые устройства, поддерживающие и MB-OFDM, и DS-UWB.

Недавно появились первые чипсеты стандарта IEEE 802.15.3. Так, Freescale Semiconductor, дочерняя компания фирмы Motorola, выпустила чипсет XS110 из трех микросхем — трансивера с ВЧ-трактом, baseband-процессора (коммуникационный процессор, выполняющий все преобразования физического уровня, включая ЦАП/АЦП) и MAC-контроллера. Максимальная скорость, обеспечиваемая модемом на основе этого чипсета, — 114 Мбит/с (т.е. речь идет о версии стандарта на основе технологии DS-UWB). Энергия, потребляемая чипсетом, — 750 мВт, напряжение питания — 3,3 В. Общая мощность излучения в полосе частот 3,1–10,6 ГГц — менее 1 мВт. Микросхемы изготовлены на основе 0,18-мкм КМОП- и SiGe-технологий.

2.1.5 Низкоскоростные сети стандарта IEEE 802.15.4 (ZigBee)

Однако мир испытывает потребность не только в сетях с высочайшими скоростями передачи данных. Для очень широкого спектра задач достаточно низкой скорости обмена — лишь бы сетевые устройства были максимально простыми, дешевыми, со сверхнизким потреблением энергии и несложным механизмом подключения к сети. Скажем, для интерактивных игр не нужна скорость обмена с компьютеров свыше 250 кбит/с, а разнообразные задачи автоматизации и системы сбора информации и вовсе не требуют скоростей передачи свыше 20 кбит/с.

Для решения данного круга задач и был разработан стандарт низкоскоростных БСПИ IEEE 802.15.4. Его разработчиком выступил альянс компаний (Inven-sys, Honeywell, Mitsubishi Electric, Motorola, Philips), назвавший себя ZigBee — (от Zig-zag — зигзаг и Bee — пчела). Подразумевалось, что топология сети бу-

дет напоминать зигзагообразную траекторию полета пчелы от цветка к цветку. Под таким замысловатым названием технология ZigBee и получает все большее распространение.

Стандарт IEEE 802.15.4 (ZigBee) предусматривает работу в трех диапазонах: один канал 868,0–868,6 МГц (для Европы), 10 каналов в диапазоне 902–928 МГц (шаг центральных частот 2 МГц, самая нижняя из них — 906 МГц) и 16 каналов в диапазоне 2400–2483,5 МГц (шаг центральных частот 5 МГц, самая нижняя из них — 2405 МГц) (табл. 2.6).

Таблица 2.6. Частотные диапазоны и скорости передачи в сетях IEEE 802.15.4

Частотный диапазон, МГц	Чиповая скорость, Кчип/с	Модуляция	Битовая скорость, кбит/с	Скорость символов, Ксимволов/с
868–868,6	300	BPSK	20	20
902–928	600	BPSK	40	40
2400–2483,5	2000	O-QPSK	250	62,5

В радиоканале использован метод широкополосной передачи с расширением спектра прямой последовательностью (DSSS). Модуляция и расширяющие последовательности для диапазонов 868/915 и 2450 МГц различны.

В диапазоне 2450 МГц поток немодулированных данных разбивается на группы по четыре бита. Каждая группа заменяется одной из 16 квазиортогональных последовательностей длиной 32 бита (чипа). Последовательности приведены в стандарте. Модуляция данных — квадратурная фазовая (QPSK). Четные чипы квазиортогональной последовательности (начиная с нулевого) модулируют синфазный (I) канал, нечетные — квадратурный (Q) канал. В результате последовательность в квадратурном канале смещена относительно синфазного на период одного чипа, поэтому модуляция называется Offset-QPSK (QPSK со сдвигом). Длительность импульса после квадратурного модулятора вдвое больше, чем длительность одного чипа (форма импульса — половина периода синусоиды с частотой, вдвое меньшей частоты чипов).

В диапазоне 868/915 МГц поток данных подвергается дифференциальному кодированию по схеме $E_i = E_{i-1} \oplus R_i$, $E_0 = 0$. Здесь R_i и E_i — биты до и после кодирования соответственно. Далее происходит замена каждого бита расширяющей последовательностью длиной 15 бит («1» заменяется на 0537₁₆, «0» — на инверсную последовательность 7AC8₁₆). Далее преобразованный поток данных передается в радиоканал посредством двухпозиционной фазовой модуляции (BPSK). Форма импульса при этом соответствует так называемому приподнятому косинусу, в данном случае — функции вида $[\sin(x)]/[x(1-x^2)]$, где $x = 2\pi f_{\text{чипов}}t$, $t = 0..1/f_{\text{чипов}}$.

Сеть стандарта IEEE 802.15.4 содержит два типа устройств — так называемые полнофункциональные (FFD) и устройства с уменьшенной функциональностью (RFD). Их основное отличие: FFD могут устанавливать соединения с любыми устройствами, RFD — только с FFD. В каждой пикосети (PAN) должно быть устройство — координатор PAN. Его функции может выполнять только FFD.

Сеть, состоящая из одного FFD и нескольких RFD, образует топологию типа «звезда». Если в сети FFD несколько, топология может быть более сложной — типа одноранговой сети (сети равноправных устройств, peer-to-peer) «каждый с каждым» (рис. 2.16) или объединение нескольких звездообразных кластеров (рис. 2.17). Но в любом случае одно из FFD выполняет функцию координатора сети. Каждому устройству сети присваивается 64-разрядный адрес. Отметим, что стандарт предусматривает взаимодействие устройств не только в рамках одной PAN, но и между различными соседними PAN (для чего и нужна развитая система адресации). Для упрощения обмена внутри сети координатор PAN может присвоить устройствам более короткие 16-разрядные адреса. В этом случае для межсетевого взаимодействия используются 16-разрядные идентификаторы сетей, также назначаемые координатором.

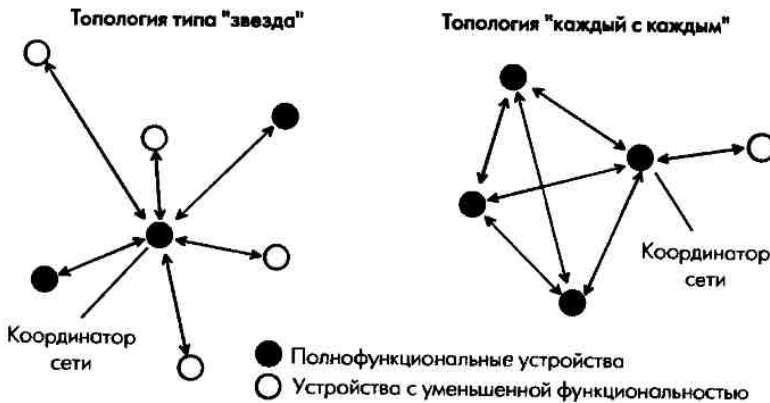


Рис. 2.16. Топология сети IEEE 802.15.4 типа «звезда» и «равный с равным»

Информационный обмен в пикосети происходит посредством последовательности суперфреймов. В общем случае суперфрейм включает управляющий интервал (beacon), за ним следует интервал конкурентного доступа (CAP) в соответствии с механизмом CSMA/CA и период назначенного доступа. Последний содержит набор временных интервалов, назначенных определенным устройствам, чувствительным к задержкам, для передачи данных (гарантированные тайм-слоты, GTS), например для связи беспроводного манипулятора «мышь» с компьютером. Управляющий интервал передает только координатор PAN. Отметим, что в суперфрейме может не быть ни управляющего интервала, ни GTS. В общем, структура суперфреймов аналогична принятой в стандарте IEEE 802.15.3 (рис. 2.14).

Каждое устройство передает информацию посредством фреймов (пакетов). Они могут быть четырех типов — управляющие (beacon frame), фреймы данных, фреймы подтверждения приема данных и фреймы команд MAC-уровня. Фреймы физического уровня (рис. 2.18) содержат заголовок с синхропоследовательностью и информацией о размере фрейма (до 127 байт) и собственно поле данных — пакет MAC-уровня. Последний содержит заголовок со всей необходимой информацией о фрейме (тип, наличие криптозащиты, необходимость подтверждения приема и т. п.), адреса и идентификаторы устройства — отправителя и полу-

чателя, собственно поле данных и проверочную контрольную сумму (рис. 2.19). Сама процедура обмена информацией может использовать пакеты подтверждения приема данных (если потеря пакета критична).

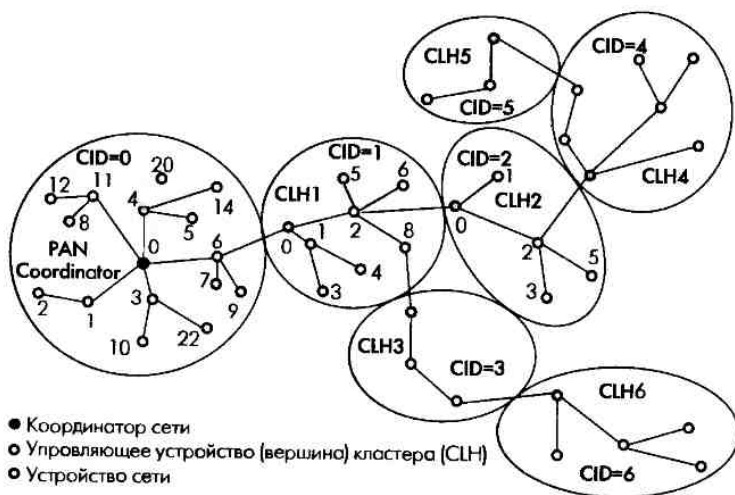


Рис. 2.17. Объединение нескольких кластеров в сети IEEE 802.15.4

Заголовок синхронизации		Заголовок физического уровня		Поле данных
Преамбула	Маркер начала фрейма	Длина фрейма	Зарезервировано	
4 байта	1 байт	7 бит	1 бит	Произвольно

Рис. 2.18. Структура пакетов физического уровня стандарта IEEE 802.15.4

Заголовок MAC-уровня						Поле данных	Проверочная последовательность
Контроль кадра	Номер последовательности	Идентификатор сети назначения	Идентификатор устройства назначения	Идентификатор сети источника	Адрес источника		
		Адресные поля					
2 байта	1 байт	0 / 2 байта	0 / 2 / 8 байт	0 / 2 байта	0 / 2 / 8 байт	Произвольно	2 байта

Рис. 2.19. Структура пакетов MAC-уровня стандарта IEEE 802.15.4

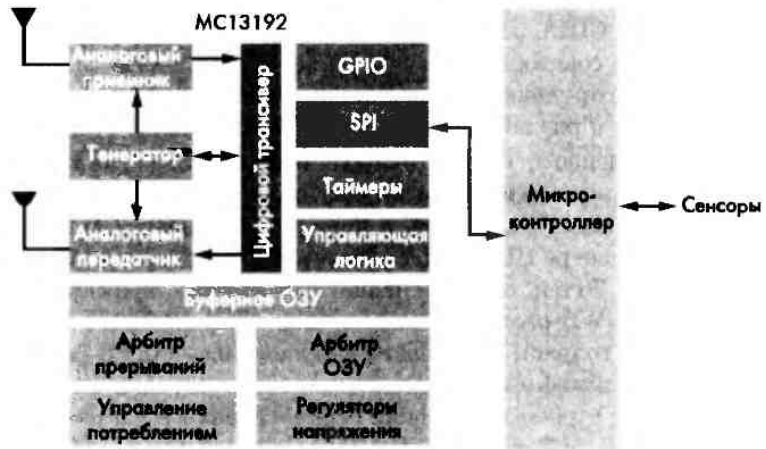
Для IEEE 802.15.4 (ZigBee) чипсеты производит уже достаточно широкий ряд производителей. Характерен продукт той же компании Freescale Semiconductor — однокристалльный модем MC13192 для диапазона 2,4 ГГц (рис. 2.20). Это — законченное решение беспроводного модема. Устройство содержит интерфейс с микроконтроллером и может применяться во множестве задач.

2.1.6 Технология сверхширокополосной связи

Рассказывая о персональных беспроводных сетях передачи данных, нельзя не остановиться на технологии сверхширокополосной связи — направлении не то что-

бы совсем новым, но обретшем в последние годы второе дыхание. Из сугубо специальной технологии для особых случаев (главным образом в военной области) оно обещает превратиться, в частности, в основу для сверхвысокоскоростных персональных БСПН.

Рис. 2.20. Однокристальный модем MC 13192 компании Freescale Semiconductor



До 1989 года для обозначения сверхширокополосных систем пользовались понятиями сигналов «без несущей», «импульсных», «несинусоидальных» и т.п. Определение термина «сверхширокополосные устройства» (UWB) введено агентством DARPA Министерства обороны США в 1990 году и скорректировано Федеральной комиссией связи США (FCC) в 2000 году. По определению FCC, к UWB-устройствам относятся все системы со спектральной полосой не менее 1,5 ГГц, все системы со спектральной полосой не менее 1,5 ГГц, а также устройства, у которых ширина спектральной полосы по уровню -10 дБ составляет по крайней мере 25% от значения центральной частоты (показатель широкополосности $\mu = 2(f_H - f_L)/(f_H + f_L) \geq 0,25$; где f_H — верхняя граница спектральной полосы по уровню -10 дБ, f_L — нижняя граница). В отечественной науке к сверхширокополосным относили сигналы, у которых ширина спектра $\Delta\omega$ соизмерима с центральной частотой ω_0 : показатель широкополосности $\mu = \Delta\omega/\omega_0 \approx 1$, в то время как у узкополосных сигналов $\mu \ll 1$ [9]. Кстати, само понятие «сверхширокополосный сигнал, система» впервые введено известным отечественным специалистом Л.Ю. Асташиным.

Работы в области сверхширокополосных систем начались в конце 50-х годов в связи с развитием радиолокационной техники как в США, так и в СССР. В США в то время несколько научных центров занимались исследованиями отклика микроволновых систем (например, антенных элементов) на короткоимпульсное (наносекунды и менее) воздействие. Такие работы в рамках создания радаров на фазированных антенных решетках вели и в исследовательском центре Sperry (позднее — Sperry Rand Corporation). В 1965 году сотрудник этого центра Джеральд Росс возглавил группу исследователей «электромагнитных явлений во временной области» (time domain electromagnetics). В результате был опубликован ряд

пионерских работ, например [192]. В 1972–1973 годах Д. Росс и другой сотрудник Sperry Rand Corporation Кеннет Роббинс получили основополагающие патенты в области сверхширокополосных систем — «Передающая и детектирующая система для генерации и приема импульсных сигналов без искажений для короткоимпульсной коммуникационной системы» [72] и «Короткоимпульсный приемник» [71]. Работы Росса и Роббинса заложили основу короткоимпульсной связи и локации в США. Достаточно сказать, что Росс является автором (самостоятельно или в соавторстве) свыше 40 патентов США в данной области. Всего же к 1989 году сотрудники Sperry Rand получили около 50 патентов США в области UWB-систем и различных аспектов их применения.

Одновременно с Россом и Роббинсом аналогичные исследования проводил Хеннинг Хармут (Американский католический университет), опубликовавший в 1969–1984 годах ряд статей и книг, посвященных UWB. В Римском исследовательском центре ВВС США работами по применению UWB-радаров занимался Паул Ван Эттен, обнаруживший ряд важных эффектов. В 1974 году Морей разработал UWB-радарную систему для исследования подземных объектов [73], что явилось толчком для нового направления — UWB-радаров для геолокации и локации «сквозь стены». Важным этапом в развитии UWB-технологий стало появление стробоскопических осциллографов с частотой сканирования выше 10 ГГц как за рубежом (Hewlett-Packard), так и в СССР. Не оставались в стороне и такие исследовательские центры, как Ливерморская и Лос-Аламосская национальные лаборатории. Отметим работы компании Multispectral Solutions, создавшей по заказу правительства США при участии Д. Росса коммуникационную систему с высокой скрытностью и помехоустойчивостью (с малой вероятностью обнаружения и противодействия, LPI/D), принятую в 1987 году. Список исследователей и работ можно продолжать. Однако к 80-м годам основные теоретические исследования в области UWB уже были завершены, созданы первые гражданские и военные системы (радиолокация, геолокация, позиционирование, связь и т. д.) [123, 126].

Аналогичные исследования велись и в СССР. Насколько можно судить по публикациям, почти все они (как и за океаном) были связаны с радарной тематикой. Российские и зарубежные эксперты отмечают работы Зернова, Карцевича, Астанина, Кобзарева, Костылева, Кардо-Сысоева, Глебовича, Мелешко и др. Кстати, первые работы по UWB системам связи с кодовым разделением каналов были выполнены в России в середине 80-х годов В.В. Крыловым в Горьковском политехническом институте. Однако с начала 90-х годов многие работы оказались если не замороженными, то сильно заторможенными.

Прорыв произошел в конце 80-х, когда в США несколько компаний очень активно начали продвигать свои UWB-системы гражданского назначения. Наиболее агрессивной из них была компания Time Domain. Ее в 1987 году основал Ларри Фуллертон, тогда же получивший свой первый патент в области сверхширокополосной связи [74]. За десять лет работы компания со своей технологией PulsOn превратилась в одного из лидеров в области портативных UWB коммуникационных связных систем и радаров, в ее активе — свыше 60 патентов по тематике UWB. У Time Domain появилось немало последователей и конкурентов [218].

Рассмотрим сущность технологии UWB на примере связной UWB-системы, описанной в патенте Л. Фуллертон «Полнодуплексная система и метод связи» [76]. Передача ведется короткими (20–0,1 нс) импульсами, следующими друг за другом с интервалом 2–5000 нс (очевидно, что значения параметров в реальных системах могут быть иными). Импульс — моноцикл Гаусса (рис. 2.21) — описывается первой производной от функции распределения Гаусса:

$$V(t) = A \frac{\sqrt{2e}}{\tau} t e^{-(t/\tau)^2}, \quad (2.1)$$

где A — амплитуда импульса, τ — временная константа, характеризующая затухание (длительность импульса $2\pi\tau$). Спектральная плотность сигнала

$$V(\omega) = A\omega\tau^2\sqrt{2\pi e}. \quad (2.2)$$

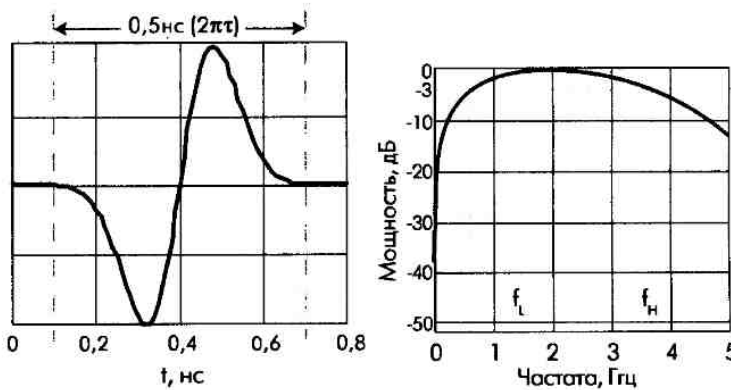


Рис. 2.21. Моноцикл Гаусса и его спектральная характеристика

Очевидно, что центральная частота такого сигнала $f_c = 1/2\pi\tau$. По уровню -3 дБ полоса сигнала ограничена частотами $f_L = 0,319f_c$ и $f_H = 1,922f_c$. Таким образом, ширина полосы сигнала в диапазоне 3 дБ составляет около 160% от центральной частоты. Для импульса длительностью 0,5 нс центральная частота равна 2 ГГц, а ширина полосы — около 3,2 ГГц. Регулярная последовательность таких импульсов не несет никакой информации, ее спектр имеет выраженный «гребенчатый» характер (рис. 2.22), следовательно, подобный сигнал может интерферировать с другими радиотехническими системами.

В «импульсном радио» (термин Фуллертон) информация кодируется посредством временной позиционно-импульсной модуляции. Смещение импульса относительно его «штатного» положения в последовательности вперед задает «0», назад — «1». Время смещения не превышает четверти длительности импульса (рис. 2.23). Так, в последовательности импульсов 0,5 нс с межимпульсным интервалом 100 нс импульс, пришедший на 100 пс раньше, — это «0», на 100 пс позже — «1». Один информационный бит кодируется последовательностью многих импульсов, например 200 импульсов на бит.

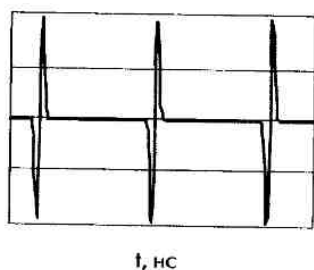


Рис. 2.22. Спектр регулярной последовательности импульсов

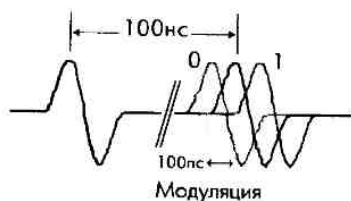


Рис. 2.23. Временная импульсно-позиционная модуляция

же полосе.

Однако возникает проблема: как разделить каналы передачи? Для этого «штатное» положение каждого импульса сдвигают на время, пропорциональное текущему значению некоторой псевдослучайной последовательности (разделение посредством временных скачков, Time Hopping). При этом время сдвига на один-два порядка выше, чем смещение при временной модуляции. В результате спектр сигнала существенно сглаживается (рис. 2.24), становится шумоподобным и уже не мешает другим устройствам, работающим в той

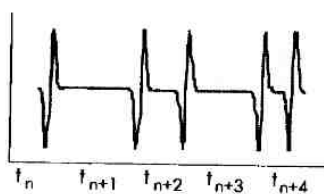


Рис. 2.24. Спектр последовательности импульсов после импульсно-кодовой модуляции

Применяя систему ортогональных кодов для управления временными задержками импульсов, теоретически можно в одной полосе создавать тысячи голосовых каналов связи [194]. По утверждению Time Domain, пессимистические оценки дают от 200 до 1000 дуплексных каналов со скоростью 64 кбит/с на одну базовую станцию без использования специальных алгоритмов цифровой обработки сигналов [188]. Таким образом, в отличие от технологии множественного доступа с кодовым разделением каналов (CDMA), системы импульсного радио используют ортогональные псевдослучайные последовательности не для расширения спектра сигнала (спектр короткого импульса и так достаточно широк), а только для сглаживания его спектральной характеристики, формирования отдельных каналов связи и защиты от помех. С другой стороны, импульсное радио можно рассматривать как предельный случай CDMA, в котором полностью отказались от несущей или, что физически почти то же самое, длительность выходного импульса сделали равной одному периоду несущей.

Кроме временного кодирования можно применять и дополнительные поднесущие — информационный сигнал предварительно модулируется тем или иным традиционным методом модуляции (AM, FM, PM, FSK, PSK, PCM и т. д.). Далее модулированные поднесущие подвергаются временной модуляции (вместо информационного сигнала модулируется поднесущая, модулированная информационным сигналом) [75].

Приемное устройство импульсного радио функционально представляет собой приемник (детектор) прямого преобразования и коррелятор. На более привычном языке его можно назвать синхронным приемником без промежуточного преобразования частоты ($f_{пч} = 0$) с системой ФАПЧ, где частота гетеродина заменяется импульсной последовательностью, позволяющей выбрать канал связи. Будучи синхронизированным с передатчиком и зная псевдослучайную последовательность канала, коррелятор определяет отклонения принятых импульсов, формируя на выходе $+1$, если сигнал, например, пришел на 100 пс раньше окончания межимпульсного интервала, -1 — если на 100 пс позже и 0 — в остальных случаях. Эти значения накапливаются в интеграторе. В результате узкополосная помеха от передатчика с непрерывной несущей или сигнал от другого импульсного передатчика способны помешать приему отдельных импульсов, но не информационного бита в целом. Накопленным значением коррелятора от случайных помех будет 0 . Разумеется, описанный метод корреляционного детектирования последовательности сверхкоротких импульсов — далеко не единственный.

Оценить помехозащищенность импульсного радио можно, используя понятие усиления обработки. В системах с расширением спектра усиление обработки определяется как отношение ширины полосы канала к ширине полосы информационного сигнала. Так, для систем расширения спектра методом прямой последовательности с шириной канала 5 МГц и информационным сигналом 10 кГц усиление составит 500 раз (27 дБ). Такой же сигнал, передаваемый импульсным радио с шириной полосы 2 ГГц, будет усилен в 200 тыс. раз (53 дБ). Если один бит задается последовательностью 200 импульсов с частотой следования 10 млн импульсов в секунду (через 100 нс) и шириной импульса 0,5 нс, скорость информационного потока составит 48,8 кбит/с, а эффективное усиление — 46 дБ. При этом усиление за счет отношения ширины межимпульсного интервала к длительности импульса ($100/0,5 = 200$) составит 23 дБ, поскольку 200 импульсов задают один бит.

Одно из существенных достоинств импульсного радио — отсутствие интерференции прямо распространяющегося сигнала с его отражениями от различных объектов. Переотражения — бич для многих связанных и локационных технологий (за исключением широкополосных шумоподобных систем связи). Именно из-за них затруднена связь внутри помещений, в условиях сложного рельефа и т. п. В импульсном радио отраженный сигнал попадет в коррелятор с задержкой и будет восприниматься как случайная помеха, никак не воздействуя на прямой сигнал. Кроме того, благодаря широкополосности затухание короткоимпульсного сигнала в различных средах достаточно мало — короткие импульсы легко проходят сквозь различные препятствия, поскольку подавление сигнала происходит не во всем диапазоне.

Из-за высокого эффективного усиления системы UWB могут работать с очень малой средней мощностью передатчика. Так, в октябре 1997 года Time Domain

демонстрировала полнодуплексную связь по технологии UWB с центральной частотой 1,7 ГГц на дистанции свыше 900 м и скоростью 32 кбит/с. При этом средняя мощность передатчика составляла около 2 мВт. Система одновременно определяла расстояние с точностью до 3 см. Разработана полнодуплексная система 1,3 ГГц передачи со скоростью 39–156 кбит/с на дистанции до 16 км. Средняя мощность ее передатчика 250 мкВт. Также компания создала демонстрационную беспроводную локальную сеть со скоростью передачи 5 Мбит/с и радиусом действия свыше 10 м (через две стены внутри здания) при мощности излучения 50 мкВт. С такой же средней мощностью работает радар «RadarVision 1000».

Средняя мощность излучения UWB-устройств может быть чрезвычайно низкой, поэтому они не должны мешать существующим радиотехническим системам, работая в одном с ними диапазоне. Поскольку UWB-сигнал распределен по столь широкому спектру, обнаружить его, а тем более перехватить или заглушить весьма проблематично. Разумеется, совокупность таких свойств, как высокая помехозащищенность, скрытность, малое энергопотребление и простота реализации, не могла не заинтересовать военные и государственные организации. Собственно, в США (равно как в СССР) практически любая деятельность в этой сфере до конца 70-х носила закрытый характер. До недавнего времени в США заказчиками UWB-систем могли выступать только правительственные и военные организации, работы имели статус правительственных программ. Ситуация изменилась в 1994 году, когда в США был снят ряд ограничений, и UWB-технологии стали стремительно развиваться. Не последнюю роль сыграла и относительная техническая простота в сочетании с возможностью интегральной реализации сверхширокополосных систем. Однако по-прежнему для использования таких систем требовалось специальное разрешение FCC, что существенно ограничивало сбыт продукции создателей UWB-систем.

Только 10 мая 2000 года FCC разрешила опытное безлицензионное использование UWB-систем. В результате для UWB-систем появились четкие нормы разработки и применения, а перспективы перед UWB-системами открываются огромные. При малой мощности UWB-системы способны передавать данные со скоростями свыше 20 Мбит/с, причем внутри помещений и объектов со сложной архитектурой. Телефония, локальные сети, системы «последней мили» и т. д. — во всех этих областях могут использоваться UWB-системы. Кроме этого, указанные системы могут широко применяться в локации, системах позиционирования, медицинской технике, транспорте и т. д. Недаром ряд крупнейших фирм, среди которых Intel и Siemens, инвестируют средства в UWB-компании. Многие производители полностью подготовились к открытию массового рынка. В частности, Time Domain разработала комплект ИС PulsOn для UWB-приложений. По утверждению Time Domain, на основе данного комплекта ИС можно создавать законченные решения систем импульсного радио стоимостью ниже 20 долл. [188].

Любая UWB-система включает антенную систему, формирующую короткие импульсы электромагнитного излучения; мощный импульсный ключ, управляющий антенной системой; устройство модуляции/демодуляции; прецизионный высокочастотный опорный генератор, приемный детектор и коррелятор (рис. 2.25). Отметим, что в UWB-системе нет мощных усилителей, ее приемник — прямого преобразования (без гетеродина и элементов частотной фильтрации), устрой-

ства детектирования и модуляции/демодуляции достаточно просты и реализуются средствами обычной цифровой логики, без сложных алгоритмов цифровой обработки сигналов. Антенные системы также достаточно просты по исполнению и могут изготавливаться непосредственно на печатных платах.

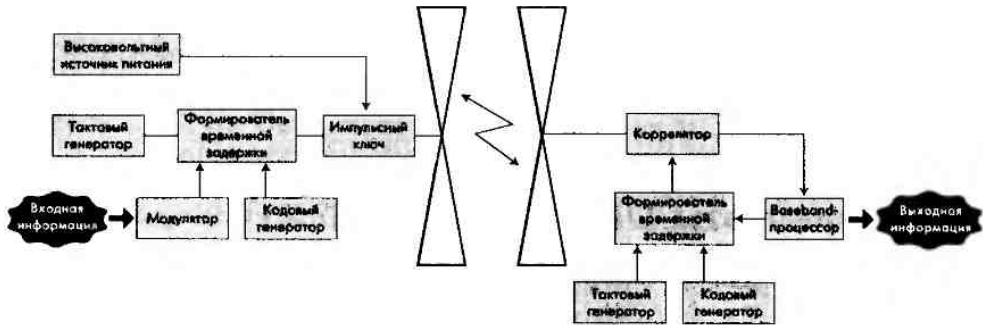


Рис. 2.25. Структура приемопередающей UWB-системы

Один из важнейших элементов для UWB-технологии — мощные импульсные ключи, особенно актуальные для локационных систем. Требования к коммутирующим элементам достаточно жесткие — они должны открываться (закрываться) на сотни пикосекунд (или, по крайней мере, иметь фронты открывания/закрывания длительностью порядка 10–100 пс) с мегагерцовой частотой повторения при очень высокой стабильности. При этом коммутируемое напряжение измеряется сотнями и тысячами вольт.

В начале 80-х годов в ФТИ им. А.Ф. Иоффе были открыты два интереснейших явления — эффекты сверхбыстрого восстановления напряжения и сверхбыстрого обратимого пробоя в высоковольтных переходах. Это открытие привело к созданию принципиально новых полупроводниковых приборов, способных коммутировать большие мощности в малых временных промежутках, в том числе приборов на задержанной ударно-ионизационной волне и дрейфовых приборов с резким восстановлением. Идея и ее материальное воплощение принадлежат группе разработчиков во главе с А.Ф. Кардо-Сысоевым, за что они были удостоены Государственной премии СССР.

На основе этих приборов создано семейство генераторов различного назначения, позволяющих формировать сверхкороткие импульсы нано- и пикосекундной длительности с мощностью до десятков мегаватт и частотой повторения до десятков мегагерц, контролируя при этом их временное положение с точностью 10 пс. Максимальное импульсное напряжение может достигать десятков киловольт, а импульсные токи — тысяч ампер. С ростом частоты повторения пиковая мощность падает, но по-прежнему остается намного выше, чем полученная с помощью любых других полупроводниковых устройств.

Это позволяет с успехом применять разработанные генераторы в локационных и связных UWB-системах. Причем ресурс этих приборов чрезвычайно высок.

Отметим, что даже сегодня вопросы массового применения UWB-технологии окончательно не разрешены. Продолжаются исследования интерференции UWB-

устройств с другими радиотехническими системами. Многие эксперты, в частности, справедливо отмечают, что стоит человеку прикоснуться к антенне UWB-передатчика, как спектр сигнала из непрерывного может стать гребенчатым — все равно что подключить режекторный фильтр. Но уже почти готов стандарт на сверхширокополосные персональные БСПИ (IEEE 802.15.3a). Практически все ведущие мировые полупроводниковые и телекоммуникационные фирмы так или иначе вовлечены в работу в области UWB-технологий, и новые результаты можно ожидать в самое ближайшее время.

2.2 Беспроводные локальные сети (стандарты DECT и IEEE 802.11)

Беспроводные локальные сети передачи информации (WLAN) развиваются в последние десять лет невероятно быстро. Простота развертывания таких сетей ограничена только необходимостью оформления разрешительной документации (в тех странах, где это требуется). По пропускной способности они не уступают выделенным медным линиям. Помехоустойчивость, надежность и защищенность современных протоколов передачи сделали WLAN явлением повсеместным, а оборудование для них — массовым продуктом. Отметим, что понятие «локальные сети передачи информации» достаточно условно. Как правило, имеются в виду системы, локализованные в радиусе сотни метров. Однако технологии локальных сетей с успехом применяют и на расстояниях до нескольких десятков километров. Современные стандарты в этой области наглядно демонстрируют, насколько тесно сливаются сегодня две телекоммуникационные ветви — традиционная телефония и системы передачи данных. В настоящем разделе кратко рассматриваются два наиболее массовых пула стандартов в области локальных сетей — IEEE 802.11 и DECT.

2.2.1 Локальные сети под управлением IEEE 802.11

Рынок массовых устройств БСПИ очень молод. Первые устройства для беспроводных локальных сетей появились в начале-середине 90-х годов. Но уже в 1999-м объем продаж устройств для беспроводных сетей достиг 600–770 млн долл., а к 2004 году он составил порядка 2,2–3 млрд долл. Ежемесячно в мире продается около миллиона адаптеров только стандарта IEEE 802.11b, по экспертным оценкам, к 2003 году их было установлено свыше 20 миллионов. Причем стремительно развиваются сами технологии передачи и оборудование для них — скорости выросли от 1–2 до 54 Мбит/с. С не меньшей стремительностью падает и стоимость оборудования.

Работы в этой области начались в 1989 году, когда была организована рабочая группа 11-го комитета IEEE 802. В июле 1997 года в результате работы этой группы был опубликован стандарт IEEE 802.11 «Спецификация физического уровня и уровня контроля доступа к каналу передачи беспроводных локальных сетей» (Wireless LAN Medium Access Control and Physical Layer Specifications). Он определял архитектуру сети и вытекающие из этого требования к функциям

устройств, принципы доступа устройств к каналам связи, формат пакетов передачи, способы аутентификации и защиты данных. Хотя стандарт изначально задумывался как инвариантный по отношению к какому-либо частотному диапазону, на физическом уровне он определял три способа работы: два радиочастотных и оптический. В инфракрасном диапазоне предусматривалась импульсно-позиционная модуляция, в диапазоне 2,400–2,4835 ГГц — режимы модуляции с расширением спектра методом частотных скачков (FHSS) и методом прямой последовательности (DSSS). Скорости обмена устанавливались на уровне 1 и 2 Мбит/с.

Отметим, что устройства, соответствующие исходной спецификации IEEE 802.11, не успели получить развития. Пропускная способность проводных сетей Ethernet сильно возросла, и максимальная скорость передачи 2 Мбит/с, предусмотренная в IEEE 802.11, не удовлетворяла пользователей. Проблему решило появление стандартов (дополнений) IEEE 802.11b, 802.11a и 802.11g.

Первым стал утвержденный 16 сентября 1999 года стандарт IEEE 802.11b. Он описывал физический и MAC-уровни беспроводных сетей для работы в диапазоне 2,4 ГГц. Стандарт определял работу на скоростях 1 и 2 Мбит/с с модуляцией только методом DSSS. Самое же главное — он предусматривал скорости обмена до 11 Мбит/с (а опционально — и до 33 Мбит/с). Передача данных на скоростях 5,5 и 11 Мбит/с происходит посредством модуляции комплементарных кодовых последовательностей CCK (основной вид модуляции). Кроме того, предусматривалась и работа на скоростях 22 и 33 Мбит/с посредством пакетного бинарного сверточного кодирования (PBCC).

Стандарт IEEE 802.11a, описывающий работу в диапазоне 5 ГГц, был принят одновременно с IEEE 802.11b. В нем использован принципиально иной, чем в IEEE 802.11b, механизм модуляции/мультиплексирования, а именно частотное мультиплексирование посредством ортогональных несущих (OFDM). Данный метод, в частности, достаточно хорошо зарекомендовал себя в системах цифрового телевизионного вещания DVB.

В IEEE 802.11a каждый пакет передается посредством 52 ортогональных несущих, каждая с шириной полосы порядка 300 кГц (20 МГц/64). Ширина одного канала 20 МГц. Несущие модулируют посредством BPSK, QPSK, 16- и 64-позиционной квадратурной амплитудной модуляции (QAM). В совокупности с различными скоростями кодирования ($1/2$ и $3/4$, для 64-QAM — $2/3$ и $3/4$) образуется набор скоростей передачи 6, 9, 12, 18, 24, 36, 48 и 54 Мбит/с. Напомним, скорость кодирования — это отношение числа бит в пакете до и после кодера. Скажем, скорость кодирования $r = 1/2$ означает, что каждый входной бит после кодирования превращается в два бита.

Диапазон 5,1–5,9 ГГц хорош тем, что там гораздо проще найти широкую полосу для системы связи. В США для безлицензионной работы в этом диапазоне выделены полосы 5,15–5,35 и 5,725–5,825 ГГц — всего 300 МГц по сравнению с 83 МГц в диапазоне 2,4 ГГц. Вместо трех неперекрывающихся каналов в диапазоне 2,4 ГГц для сетей IEEE 802.11b только в нижнем поддиапазоне 5,15–5,35 ГГц имеются восемь неперекрывающихся каналов (рис. 2.26). Аналогичная ситуация в Европе и в России (однако в нашей стране отсутствуют безлицензионные диапазоны) — в более высокочастотной области места больше. В частности, если в

Москве диапазон 2,4 ГГц занят операторами достаточно давно, то область 5 ГГц еще только начинают осваивать, хотя свободных поддиапазонов там уже практически нет.

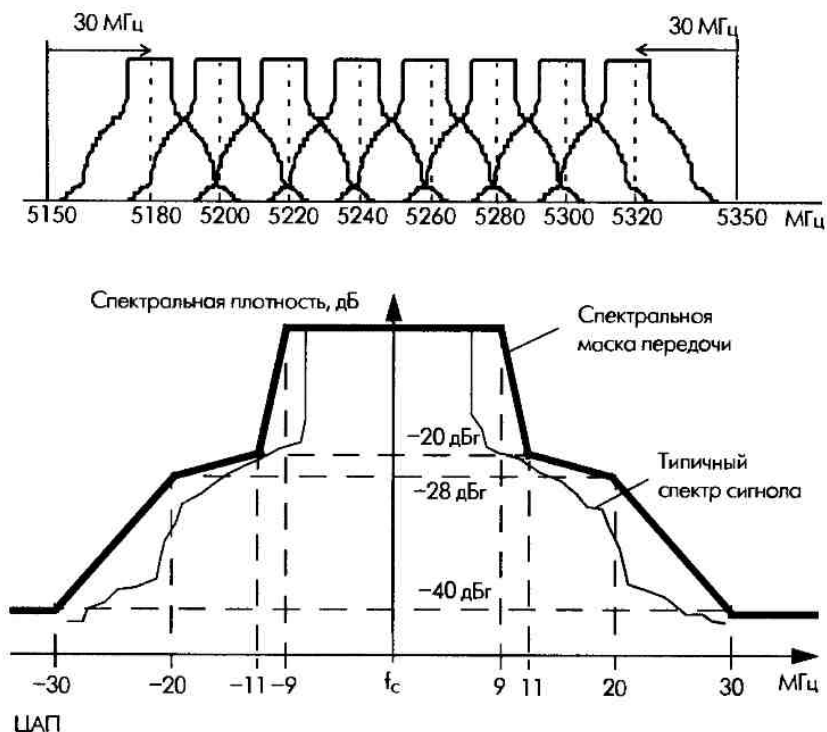


Рис. 2.26. Спектральная маска и распределение неперекрывающихся каналов в диапазоне 5,15–5,35 ГГц для стандарта IEEE 802.11a. Мощность (дБм) определяется относительно пиков функции $\sin(x)/x$

Кроме того, диапазон 2,4 ГГц перегружен различными системами — тут и беспроводные телефоны, и устройства Bluetooth (IEEE 802.15.1), и многочисленное оборудование стандарта IEEE 802.11b. Взаимных помех избежать тяжело. Сравнительные испытания убедительно показывают, что в одних и тех же условиях устройства IEEE 802.11a по скорости обмена превосходят оборудование IEEE 802.11b (рис. 2.27). И до недавнего времени казалось, что будущее принадлежит сетям стандарта IEEE 802.11a. Однако возник ряд вопросов.

Прежде всего, как быть с уже существующими сетями (и оборудованием) в диапазоне 2,4 ГГц? Как обеспечить столь необходимую всем обратную совместимость? С этой проблемой производители справились, разработав двухдиапазонные чипсеты. Характерный пример — компания Atheros, создавшая комплект из трех ИС AR5001X Combo. В его составе ИС baseband-процессора и MAC-контроллера AR5211 и две аналоговые ИС трансиверов — на 2,4 и на 5 ГГц (AR2111 и AR5111 соответственно). Аналогичное решение предложено и компанией Intersil — чипсет PRISM Duette из двух ИС (рис. 2.28): baseband/MAC-процессор ISL3890 и однокристалльный двухдиапазонный трансивер ISL3690.

Однако к моменту, когда данные решения стали технологически возможны и рентабельны, в Европе был разработан свой стандарт 5 ГГц — HiperLan2. Кроме того, работы по ускорению стандарта IEEE 802.11b в диапазоне 2,4 ГГц привели к появлению новой версии — IEEE 802.11g, предусматривающей скорости до 54 Мбит/с. В частности, рассмотренные выше чипсеты AR5001X Combo и PRISM Duette в диапазоне 2,4 ГГц поддерживают не только IEEE 802.11b, но и IEEE 802.11g.

Работы над будущей спецификацией IEEE 802.11g начались в марте 2000 года, когда была сформирована исследовательская группа по изучению возможности увеличения скорости передачи данных свыше 20 Мбит/с в диапазоне 2,4 ГГц. В ноябре 2000 года эта группа приобрела статус штатной группы разработчиков и получила обозначение G. Через полтора года, рассмотрев несколько альтернативных подходов, специалисты исследовательской группы G предложили использовать применяющуюся в стандарте IEEE 802.11a систему кодирования с мультиплексированием посредством ортогональных несущих OFDM. В качестве дополнительных (необязательных) возможностей новый стандарт IEEE 802.11g предусматривал использование таких схем модуляции, как CCK-OFDM и RBCC.

Новая спецификация по сути представляет собой перенесение схемы модуляции OFDM, прекрасно зарекомендовавшей себя в 802.11a, из диапазона 5 ГГц в область 2,4 ГГц при сохранении возможностей устройств стандарта 802.11b. Это возможно, поскольку в стандартах 802.11 ширина одного канала в диапазоне 2,4 и 5 ГГц схожа — 22 МГц по уровню -30 и -20 дБ соответственно. Правда, по уровню -28 дБ маска канала в IEEE 802.11a допускает спектральную полосу шириной 40 МГц, что может создать проблемы, безусловно, преодолимые.

Зачем нужен новый стандарт в диапазоне 2,4 ГГц, когда оборудование 802.11a для диапазона 5 ГГц становится практически массовым? Однозначного ответа здесь нет, укажем лишь на основные резоны. Прежде всего, оборудование 2,4 ГГц должно быть дешевле. Если сравнивать стоимость аналогичного оборудования одной компании стандартов 802.11a и b, цены в первом случае окажутся в среднем в 1,5–2 раза выше. Однако на уровне интегральных схем различие не столь значительное и, безусловно, имеет тенденцию к снижению. Кроме того, затухание сигналов в диапазоне 5 ГГц выше, чем в диапазоне 2,4 ГГц. Поэтому плотность оборудования в сетях 5 ГГц тоже должна быть выше, вместе с ценой сети. Конечно, минус этот относительный, поскольку в ряде случаев зону распространения сигналов оборудования локальной сети даже желательно ограничить,

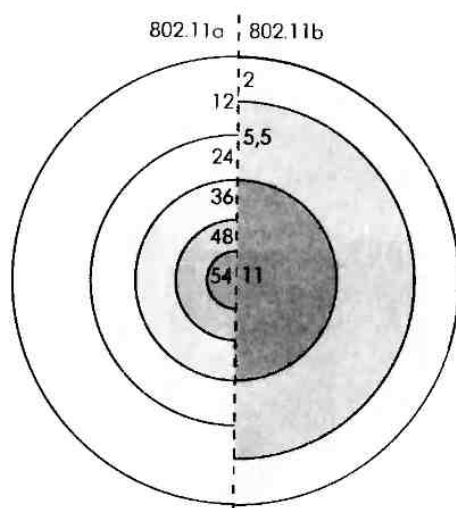


Рис. 2.27. Соотношение скоростей передачи устройств стандартов 802.11a и 802.11b (по материалам компании Proxim)

например из соображений защиты информации или чтобы избежать взаимного влияния двух разных сетей.

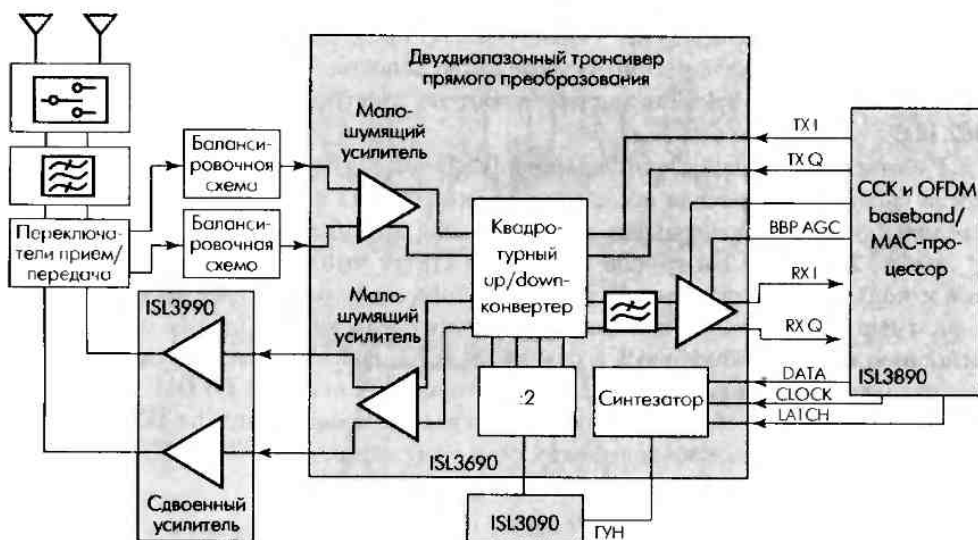


Рис. 2.28. Чипсет PRISM Duette

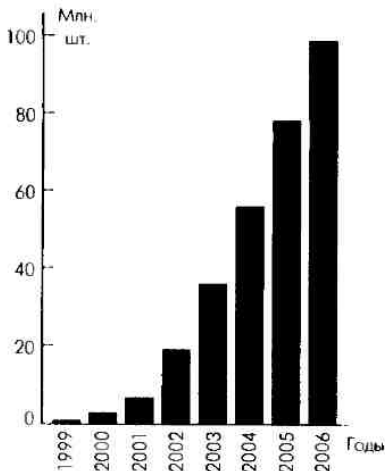


Рис. 2.29. Динамика рынка устройств для сетей IEEE 802.11. По материалам компании IC Insights

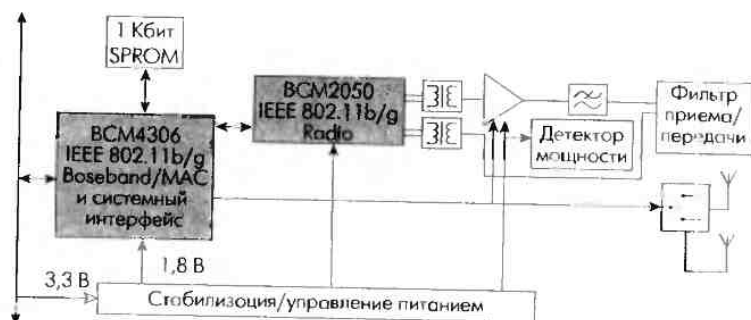
Устройства IEEE 802.11g с 2002 года производят такие компании, как Buffalo Technologies, Linksys, D-Link, Apple. Позднее к ним присоединились фирмы Netgear, Belkin, Actiontec, Proxim и многие другие. Такую возможность им предоставили производители наборов микросхем для 802.11g (прежде всего компании Intersil, Atheros Communications, Broadcom). Лидером в данной области выступает компания Intersil, продавшая в 2002 году компонентов для WLAN на 106 млн долл. Однако наиболее динамичной корпорацией в этой области выступает Broadcom — первый свой чипсет для 802.11b она представила в июле 2002 года, а в ноябре уже объявила о серийном производстве чипсетов 802.11g. К марту 2003 года Broadcom продала 1,3 млн чипсетов.

Столь высокие темпы освоения производства изделий IEEE 802.11g продиктованы быстрым ростом рынка этих устройств (рис. 2.29). В 2002 году продано, по разным оценкам, 18–19 млн чипсетов для беспроводных сетей (практически все — стандарта IEEE 802.11b с 5%-ым вкраплением 802.11a), в 2003 году аналитики компании Forward Concepts прогнозировали объем продаж на уровне 30 млн комплектов ИС (на 70% больше!) [121]. Эксперты компании IC

Insights предсказывают, что если к концу 2003 года объем рынка оборудования для беспроводных сетей составит 4,1 млрд долл., то к 2006 году он вырастет до 9,3 млрд! Стремительно падает и цена чипсетов — она уже ниже 10 долл. Несмотря на это объем продаж ИС для беспроводных сетей возрастет: если в 2002 году он составлял 471 млн долл., к 2006 году он может превысить миллиардную отметку.

Первым набором микросхем для устройств IEEE 802.11g компании Broadcom стал чипсет для абонентских устройств BCM94306. Он входит в состав большого семейства продуктов компании для беспроводных сетей AirForce. Чипсет включает две ИС — baseband/MAC процессор BCM4306 и однокристалльный радиомодуль BCM2050 (рис. 2.30).

Рис. 2.30. Построение абонентского устройства 802.11g на базе чипсета BCM94306



Процессор поддерживает интерфейс к шинам PCI/PCMCIA, а также асинхронный последовательный интерфейс для коммутации с микросхемами, обеспечивающие связь посредством протоколов Bluetooth и GPRS (у компании Broadcom есть для этого однокристалльные решения). Отметим, что в семейство AirForce входит широкий спектр продуктов — как абонентские устройства (рис. 2.31) различного исполнения (CardBus, Mini PCI, USB и др.), так и точки доступа маршрутизаторы с поддержкой высокоскоростных проводных интерфейсов [117].

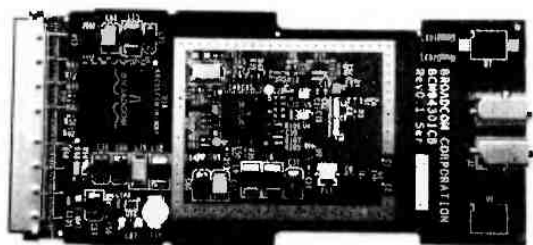


Рис. 2.31. Сетевая абонентская карта BCM94306CB стандарта 802.11b/g в формате CardBus компании Broadcom

Практически одновременно с компанией Broadcom к производству микросхем для устройств 802.11g приступили корпорации Intersil и Atheros Communications. Intersil выпустила чипсеты PRISM Duette (802.11a/b/g) и PRISM GT (802.11b/g). Компания Atheros представила на рынок чипсет AR5001X Combo с поддержкой 802.11a/b/g. В марте 2002 года это было первое промышленное решение «три в одном» (рис. 2.32), поскольку PRISM Duette компания Intersil анонсировала в октябре 2002 года.



Рис. 2.32. Чипсет AR5001X Combo с поддержкой 802.11a/b/g компании Atheros

Предполагается, что по мере снижения стоимости высокочастотных компонентов доминирующее положение на рынке займут двухдиапазонные, устройства 2,4 и 5 ГГц стандарта IEEE 802.11a/b/g. Аналитики компании IC Insights предсказывают, что доля таких устройств в общем объеме произведенных в 2006 году (97 млн) составит 68% (рис. 2.33) [177]. По прогнозу же фирмы Forward Concepts, в 2006 году устройства, поддерживающие только 802.11a, исчезнут с рынка [121].

Еще одно важное направление развития — это совершенствование защиты информации и качества сервиса (QoS) в сетях, над чем работают исследовательские группы IEEE 802.11i и 802.11e соответственно. Если удастся решить эти вопросы, сети 802.11 могут стать универсальной платформой, на которой будут развиваться беспроводные услуги, включая мобильную связь. Ожидается и официальное утверждение нового высокоскоростного стандарта WLAN со скоростями обмена выше 100 Мбит/с — IEEE 802.11n.



Рис. 2.33. Состояние и прогноз рынка устройств группы стандартов 802.11. По материалам компании IC Insights

2.2.2 Стандарт DECT

Назначение стандарта DECT

DECT — стандарт уникальный. Задуманный для телефонии, он практически сразу стал использоваться в системах передачи данных. Изначально предназначенный для Европы, распространился по всему миру. DECT конкурирует со стандартами сотовой связи, радиорелейными технологиями, проникает в домашние мультимедийные системы, становится средством первичного доступа в общественные телефонные сети, входит в пул стандартов сотовой телефонии третьего

поколения IMT-2000. Рынок систем DECT до сих пор остается одним из наиболее динамичных в мире.

Исторически DECT был призван избавить пользователей телефонных аппаратов от соединительных шнуров. С развитием интегральных полупроводниковых технологий телефоны стали оснащать приемопередающими блоками: появились первые радиоудлинители — обычные аналоговые телефоны, в которых шнур заменен радиотрактом. Это поколение телефонов называют CT-0 (Cordless Telephone). Такие устройства существуют и поныне. Их основное назначение — позволить владельцу свободно перемещаться в радиусе десятков-сотен метров от точки подключения к телефонной сети. Главные недостатки данных устройств — относительно высокая мощность излучения (до 1 Вт), взаимные помехи, абсолютная открытость для подслушивания и несанкционированного подключения к радиотракту. В 80-х годах в Европе появились системы стандарта CT-1 — те же аналоговые радиоудлинители, но с зачатками функций современных беспроводных средств связи, такими как роуминг и перемещение между сотами без разрыва соединения.

Таблица 2.7. Основные характеристики систем связи стандартов CT-2 (Tangara RD), CT3 и DECT

Параметр	CT-2 Tangara RD	CT-3	DECT
Диапазон рабочих частот, МГц	864–868,2	862–866	1880–1900
Способ разделения каналов	FDMA	TDMA с многочастотным разделением (4 канала), 8 дуплексных каналов на несущую	TDMA с многочастотным разделением (10 каналов), 12 дуплексных каналов на несущую
Ширина полосы канала, МГц	0,1	1	1,728
Организация дуплексной связи	TDD	TDD	TDD
Кодирование речи	ADPCM, 32 Кбит/с	—	ADPCM, 32 Кбит/с
Число каналов, поддерживаемых базовой станцией	2–6	32	до 120
Мощность передатчиков, мВт	10	—	10–250

Однако подлинным прорывом стало появление цифровой спецификации CT-2 (табл. 2.7). Разработана эта спецификация в Великобритании в 1989 году (общий беспроводной стык Common Air Interface, CAI/CT-2, стандарт MPT 1375).

В 1992 году ETSI принял CAI/CT-2 в качестве европейского стандарта. На основе CT-2 была создана система Telepoint, получившая достаточно широкое распространение во всем мире. В ней фактически впервые была реализована микросотовая архитектура (рис. 2.34). Системам CT-2 предрекали большое будущее, однако после резкого снижения цен на услуги сотовой связи во второй половине 90-х годов прошлого века интерес к ним унал [32]. В начале 90-х появились системы на основе стандарта CT-3 фирмы Ericsson. Они были установлены в ряде стран, однако вскоре внимание европейского телекоммуникационного сообщества переключилось на новую спецификацию, названную Digital European Cordless Telecommunications — DECT.



Рис. 2.34. Микросотовая сеть связи

В 1988 году новый стандарт был одобрен Конференцией европейских администраций почты и электросвязи (СЕПТ). Под него был выделен диапазон 1880–1900 МГц. В 1992 году ETSI публикует первые спецификации DECT — ETS 300 175 и 176 (DECT Common Interface (CI) и DECT Approval Test Specification соответственно). В этих документах были заложены базовые основы DECT.

Фактически DECT — это набор спецификаций, определяющих радиointерфейсы для различных видов сетей связи и оборудования. DECT CI содержит требования, протоколы и сообщения, обеспечивающие взаимодействие сетей связи и оконечного оборудования. Организация самих сетей и устройство оборудования в стандарт не входят.

Важнейшая задача DECT — обеспечить совместимость оборудования различных изготовителей. Для этого был разработан ряд профилей взаимодействия

различных систем. В 1994 году появился первый из них, унифицированный профиль доступа GAP (Generic Access Profile) — ETS 300 444. Он определяет работу оконечных устройств DECT (телефоны, базовые станции, беспроводные офисные АТС) для всех приложений голосовой связи с полосой пропускания речевого тракта 3,1 кГц. Позднее появились профили взаимодействия DECT и GSM, DECT и ISDN, взаимодействия абонентов с ограниченной мобильностью с сетями общего пользования (Cordless Terminal Mobility, CTM), со средствами абонентского радиодоступа (Radio Local Loop, RLL) и т. д.

Первые коммерческие продукты в стандарте DECT были представлены на выставке CeBIT'93 в Ганновере. Это были беспроводной телефон Gigaset производства Siemens, радио-АТС Freeset (DCT-1800) фирмы Ericsson и радиосистема для передачи данных NET3 компании Olivetty. Судьба их сложилась по-разному — о продукте NET3 помнят разве что специалисты, в то время как семейство Gigaset активно производится и развивается до сих пор. На CeBIT'96 уже пять ведущих европейских производителей, Alcatel, Ericsson, Nokia, Philips и Siemens — демонстрировали свои изделия для DECT. Тогда же DECT приказом Минсвязи был легализован в России.

Изначально DECT был ориентирован на телефонию — радиоудлинители, беспроводные учрежденческие АТС, предоставление радиодоступа к телефонным сетям общего пользования. Но стандарт оказался столь удачным, что его стали использовать в системах передачи данных, беспроводного абонентского доступа к сетям связи общего пользования. DECT нашел применение в приложениях мультимедиа и домашних радиосетях, для доступа в Интернет и к факсимильной связи.

Бурно прогрессирует и технология беспроводного доступа посредством радиоретрансляционных DECT-систем. Уже к началу 1997 года число таких линий превысило 1 млн. По данным компании Strategy Analytics, в 1998 году на европейском рынке было продано около 18 млн беспроводных телефонов DECT. По данным на январь 1999 года, 53% всех продаваемых в Европе телефонов с радиоудлинителем были стандарта DECT, причем в Германии этот показатель составил 80%.

Системы и устройства DECT распространены во всех странах мира. На мировом рынке представлены более 200 различных продуктов DECT [163]. Не случайно сейчас аббревиатура DECT расшифровывается как Digital Enhanced (вместо European) Cordless Telecommunication.

Принципы организации DECT

Что же представляет собой радиоинтерфейс DECT? В диапазоне шириной 20 МГц (1880–1900 МГц) выделено 10 несущих частот с интервалом 1,728 МГц. В DECT применяется технология доступа с временным разделением каналов — TDMA (Time Division Multiple Access). Временной спектр разделен на отдельные кадры по 10 мс (рис. 2.35). Каждый кадр разбит на 24 временных слота: 12 слотов для приема (с точки зрения носимого терминала) и 12 — для передачи. Таким образом, на каждой из 10 несущих частот формируется 12 дуплексных каналов — всего 120. Дуплекс обеспечивается временным разделением (с интервалом 5

мс) приема/передачи (TDD, Time Division Duplex). Для синхронизации применяется 32-битная последовательность «101010...». В DECT предусмотрено сжатие речи в соответствии с технологией адаптивной дифференциальной импульсно-кодовой модуляции АДИКМ (ADPCM) со скоростью 32 кбит/с (рекомендация ITU-T G.726). Поэтому информационная часть каждого слота — 320 бит. При передаче данных возможно объединение временных слотов. В радиотракте используется частотная модуляция с фильтром Гаусса (GFSK).

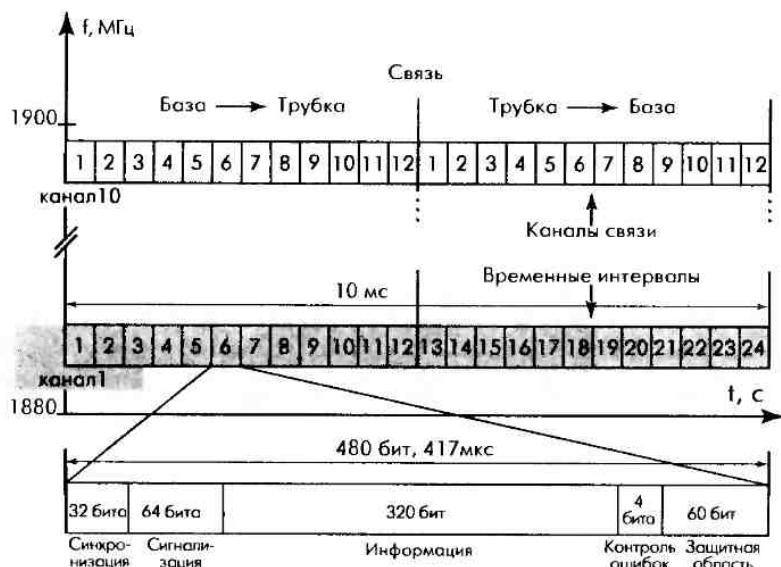


Рис. 2.35. Передача информации в системе DECT

Базовые станции (БС) и абонентские терминалы (АТ) DECT постоянно сканируют все доступные каналы (до 120). При этом измеряется мощность сигнала на каждом из каналов, которая заносится в список RSSI (Received Signal Strength Indication). Если канал занят или сильно зашумлен (например, помехами от другого DECT-устройства), показатель RSSI для него высокий. БС выбирает канал с самым низким значением RSSI для постоянной передачи служебной информации о вызовах абонентов, идентификаторе станции, возможностях системы и т. д. Эта информация играет роль опорных сигналов для АТ — по ним абонентское устройство определяет, есть ли у него право доступа к той или иной БС, предоставляет ли она требуемые абоненту услуги, есть ли в системе свободная смкость, и выбирает базовую станцию с наиболее качественным сигналом.

В DECT канал связи всегда выбирает АТ. При запросе соединения от базовой станции (входящее соединение) АТ получает уведомление и выбирает радиоканал. Служебная информация передается БС и анализируется АТ постоянно, следовательно АТ всегда синхронизируется с самой близкой из доступных БС. При установлении нового соединения АТ выбирает канал с самым низким значением RSSI — это гарантирует, что новое соединение происходит на самом «чистом» канале из доступных. Данная процедура динамического распределения каналов

DCS (Dinamic Channel Selection) позволяет избавиться от частотного планирования — важнейшее свойство DECT.

Поскольку АТ постоянно (даже при установленном соединении) анализирует доступные каналы, может происходить их динамическое переключение во время сеанса связи. Такое переключение возможно как на другой канал той же БС, так и на другую БС. Эта процедура называется «хэндовер» (handover). При хэндовере АТ устанавливает новое соединение, и какое-то время связь поддерживается по обоим каналам. Затем выбирается лучший. Автоматическое переключение между каналами разных БС происходит практически незаметно для пользователя и полностью инициируется АТ. Это особенно важно для построения микросотовых систем, позволяющих абоненту переходить из соты в соту без прерывания соединения. Отметим, что, хотя выбор каналов всегда за АТ, в DECT предусмотрена возможность оповещения абонентского терминала со стороны БС о низком качестве связи, что может инициировать хэндовер.

Существенно, что в радиотракте аппаратуры DECT мощность сигнала весьма мала — от 10 до 250 мВт. Причем 10 мВт — практически номинальная мощность для микросотовых систем с радиусом соты 30–50 м внутри здания и до 300–400 м на открытом пространстве. Передатчики мощностью до 250 мВт используют для радиопокрытия больших территорий (до 5 км при направленной антенне). Столь низкая мощность делает устройства DECT наиболее безопасными для здоровья. Недаром в европейских медицинских учреждениях разрешено применение систем радиотелефонии только этого стандарта.

Кроме того, при мощности 10 мВт возможно располагать базовые станции на расстоянии 25 м. В результате достигается рекордная плотность одновременных соединений — до 10000 Эрл./км² (около 100 тыс. абонентов) при условии расположения БС по схеме шестиугольника в одной плоскости (на одном этаже). Это лучший показатель и с точки зрения эффективности использования радиоспектра (в пересчете на 1 МГц полосы). Сравните — 500 Эрл./МГц/км² для DECT против 100 Эрл./МГц/км² в наиболее емких сотовых сетях GSM-1800 (DCS 1800).

Защита от несанкционированного доступа

Основная проблема любых систем с радиодоступом — обеспечить защиту от несанкционированного подключения и прослушивания. Когда аппаратура DECT используется как средство доступа к телефонным сетям общего пользования, прежде всего возникает опасность появления «двойников» зарегистрированных АТ. В самом деле, защита обычных аналоговых телефонов с радиоудлинителем от несанкционированного подключения давно стала постоянной головной болью для их владельцев [5]. В системах DECT данная проблема решается посредством процедур аутентификации БС и АТ.

В простейшем случае каждый АТ регистрируется в системе или на отдельных БС, к которым имеет доступ. Если речь идет о домашнем беспроводном телефоне, АТ (трубка) зарегистрирован на одной БС. При каждом соединении происходит аутентификация трубки: БС посылает АТ «запрос» — случайное число (64 бит). АТ и БС на основании этого числа и ключа аутентификации по заданному алгоритму вычисляют аутентификационный ответ (32 бит), ко-

торый АТ передает на БС. БС сравнивает вычисленный аутентификационный ответ с принятым и при их совпадении разрешает подключение АТ. В DECT существует стандартный аутентификационный алгоритм DSAA (DECT Standard Authentication Algorithm). Он поставляется производителям только по контракту с ETSI, однако даже его знание делает несанкционированное подключение весьма проблематичным.

Как правило, ключ аутентификации вычисляется на основании абонентского аутентификационного ключа UAK (User Authentication Key) длиной 128 бит либо аутентификационного кода AC (16–32 бита). UAK хранится в ПЗУ АТ либо в карточке DAM (DECT Authentication Module) — аналоге SIM-карты в стандарте GSM. AC можно записать в ПЗУ АТ и вручную либо вводить его при аутентификации. Совместно с UAK применяют и персональный идентификатор пользователя UPI (User Personal Identity) длиной 16–32 бита, вводимый только вручную. Существует ряд других процедур и механизмов аутентификации и шифрования информации в сетях DECT [54]. Кроме того, заметим, что несанкционированный съем информации в системах с TDMA доступен только специалистам.

Системы DECT

Пожалуй, самое важное в DECT то, что он — динамично развивающийся стандарт. Сочетание различных профилей и протоколов DECT позволяет создавать системы совершенно разного назначения и стоимости — от бытовых беспроводных телефонов и мини-АТС до защищенных систем передачи данных и предоставления абонентского доступа к общественным телефонным сетям. Гибкая совместимость с такими стандартами, как ISDN и GSM, делает DECT мощным средством расширения абонентской емкости существующих сетей. Особенно ярко это проявляется по отношению к сетям сотовой связи стандарта GSM. Поскольку на физическом уровне протоколы DECT и GSM достаточно близки (частотный диапазон, технология TDMA), ряд фирм выпускают сотовые GSM-телефоны с поддержкой режима DECT. Когда такой телефон находится в зоне действия БС DECT, он работает в режиме DECT, когда связь с БС ухудшается, он переключается на работу с сетью GSM.

Аппаратура DECT существенно дешевле сотовых систем, что позволяет строить коммерческие сети связи для абонентов с ограниченной мобильностью с тарифами, в два-три раза более низкими, чем в сотовой телефонии. Данный стандарт все чаще применяют для телефонизации городских и сельских районов, где кабельные системы по каким-либо причинам неприемлемы. Это происходит и в России. Например, сети абонентского радиодоступа на базе аппаратуры Goodwin WLL (компания Goodwin Europe, Россия) уже развернуты в ряде районов Московской области (Куркино, Балашиха, Реутово), в Калининграде, Твери, Таганроге и т. д.

Примером системы беспроводной связи является система Freeset (DCT-1800) фирмы Ericsson (напомним, она была первой DECT-системой связи на рынке). Freeset позволяла устанавливать входящие и исходящие соединения, передавать голосовые и факсимильные сообщения, видеoinформацию, данные, предоставлять услуги ISDN. Она состоит из контроллеров, базовых станций и абонентских

терминалов. Один контроллер управляет 80 базовыми станциями. Он подключается к телефонным сетям общего пользования посредством шин E1 (G.703, 2 Мбит/с) или двухпроводных аналоговых абонентских линий и направляет компрессированные речевые сигналы (32 кбит/с, ADPCM) базовым станциям. Система может поддерживать до 600 абонентов, максимальное удаление БС от контроллера — 2,7 км, радиус зоны обслуживания базовой станции — до 200 м. Основное назначение системы — обеспечить персональной беспроводной связью пользователей на небольшой территории (крупное здание, стройплощадка, стадион и т. д.).

Другая система Ericsson — DRA1900 — предназначена для предоставления беспроводного абонентского доступа (WLL) на существенно больших площадях, чем DCT-1800. К телефонным сетям общего пользования она подключается через ATC AXE-10 (Ericsson) посредством узлового контроллера. Узел доступа системы DRA1900 — объединение шести базовых станций, связанных с узловым контроллером каналом E1. Радиус действия БС — до 5 км. Система DRA обладает значительными возможностями гибкого расширения в зависимости от плотности абонентов и необходимой площади радиопокрытия. Системы DCT-1800 и DRA1900 сертифицированы и в России (сертификаты № ОС/1-РС-859 и ОС/1-РС-860 соответственно). DRA1900 работает в ряде регионов нашей страны, например в Нижегородской области и Тольятти.

Однако огромный размах приобрело производство беспроводных телефонов DECT. Уже в 1999 году на выставке CeBIT'99 эти изделия демонстрировали свыше 40 фирм-производителей, одна из которых — российская компания Goodwin.

Аппаратная реализация DECT-устройств

Динамичный рынок аппаратуры DECT (растущий на 30% в год) не может обойтись без соответствующей поддержки полупроводниковой индустрии. Действительно, многие производители успели выпустить несколько «поколений» комплектов ИС для DECT. Интегральные компоненты для DECT производят такие известные фирмы, как Analog Devices, GaAsTEK, Hewlett-Packard, Infineon (так стало называться подразделение полупроводников фирмы Siemens), Level One, National Semiconductor, NEC, Philips Semiconductor, ST Microelectronics, TEMIC Semiconductor, Texas Instruments, VLSI Technology и др.

Рассмотрим устройство DECT-систем на базе комплекта ИС фирмы Philips Semiconductor. Структура телефона-трубки и базовой части принципиально различается только управляющими контроллерами и наличием интерфейса к телефонной линии у базовой станции (рис. 2.36). Интерфейс может быть как цифровым, так и аналоговым. Для DECT-приложений Philips выпускает семейство микроконтроллеров PCD509х: PCD5091 — для телефонов, PCD5092–PCD5095 — для базовых станций и PCD5096 — универсальный кодек. Радиотракт формируют высокочастотный трансиверный модуль (включает супергетеродинный приемник с двойным преобразованием частоты) с подавлением помех по зеркальному каналу UAA2067 (рис. 2.37), схема гауссовой частотной модуляции на промежуточной частоте SA639 и двухчастотный синтезатор UMA1022.

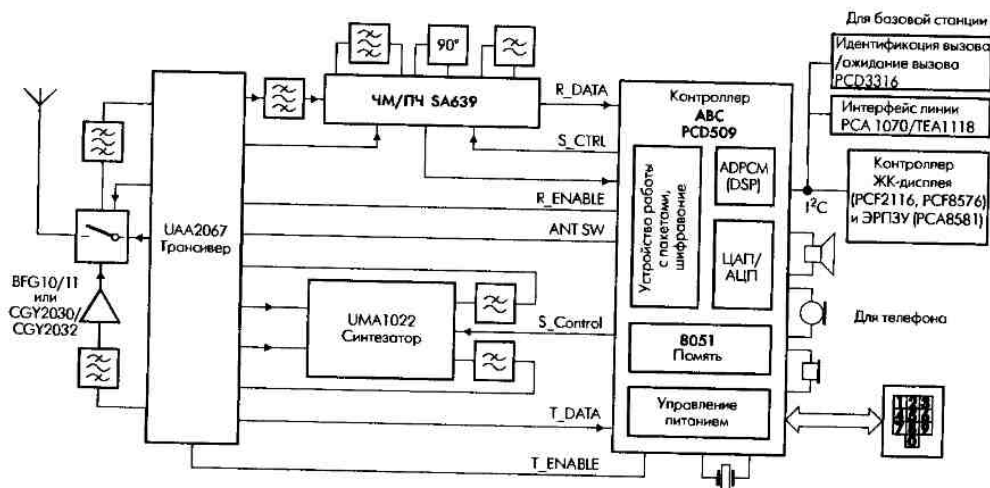


Рис. 2.36. Структурная схема телефона/базовой станции на элементной базе фирмы Philips Semiconductor

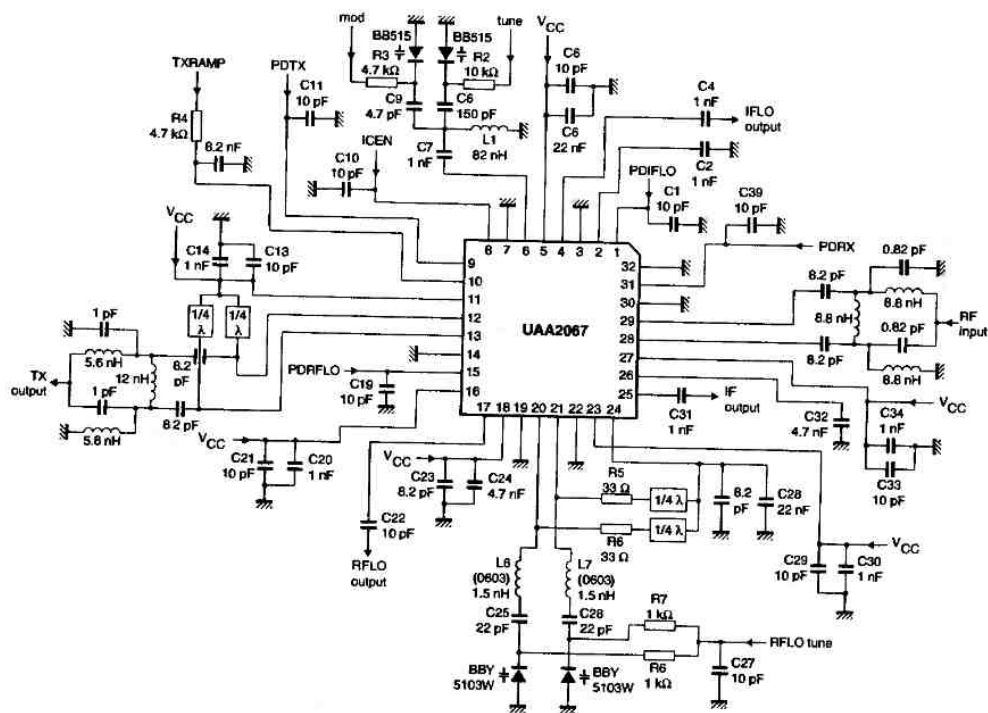


Рис. 2.37. Пример схемы включения трансивера UAA2067

ДЕСТ-контроллеры семейства РСх509х (рис. 2.38) представляют собой специализированные схемы на базе управляющего контроллера 80с51 со встроенной

(для домашних систем), PMB 5722 и 5723 (для бизнес-систем), контроллеры базовых станций PMB 5725, PMB 5735 (домашние системы) и PMB 4729 для малых офисов. Радиотракт поддерживают ИС передатчика с ФАПЧ PMB 4220, универсальный приемник PMB 4420 и кремниевый усилитель для DECT PMB 4820.

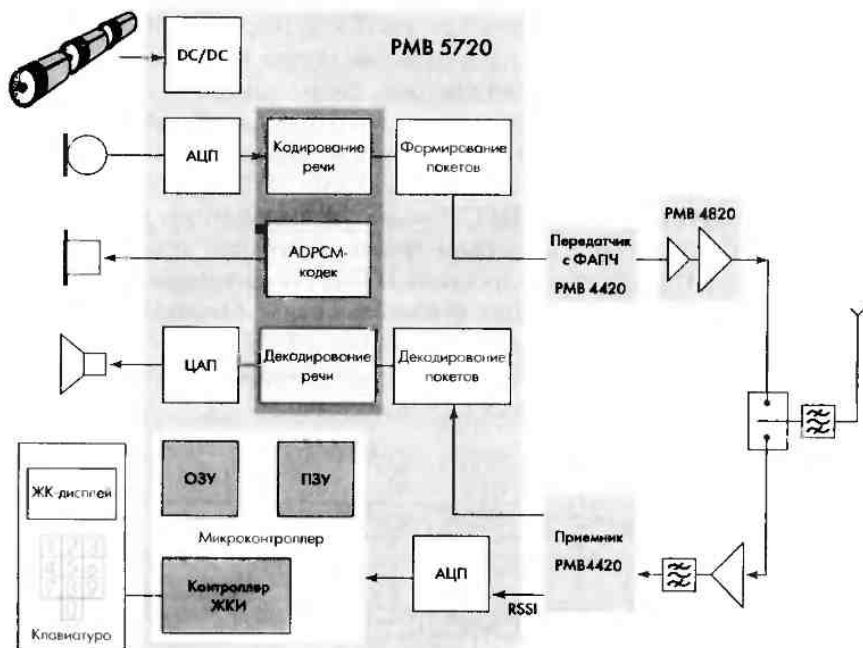


Рис. 2.39. Построение телефона DECT на базе комплекта ИС фирмы Infineon

Мировая микроэлектронная индустрия готова предоставить изготовителям самые современные компоненты для построения как простых и дешевых устройств, так и столь угодно сложных и многофункциональных DECT-систем. Тем более что сам стандарт DECT постоянно пополняется новыми возможностями, следовательно, расширяется область его применения, увеличивается разнообразие создаваемых на его базе устройств. Так, в 1998–1999 годах в него вошли два новых дополнения — стандарт пакетной радиопередачи данных DPSR (DECT Packet Radio Service) и основанный на нем мультимедийный протокол DMAP (DECT Multimedia Access Profile). Устройства, поддерживающие эти стандарты, уже представлены на рынке (например, реализующие доступ в Internet по каналам DECT). DECT осваивает новые частотные диапазоны, растут скорости передачи данных. Он входит в пул стандартов систем мобильной связи третьего поколения IMT-2000 и продолжает бурно развиваться.

2.3 Беспроводные сети регионального масштаба

Региональные (городские) широкополосные сети используют частотные диапазоны выше 1 ГГц (включая миллиметровые диапазоны), широкие рабочие полосы

частот, высокие скорости модуляции, мощные выходные усилители. Ранее это ассоциировалось с тяжелым и сложным оборудованием. Однако успехи микроэлектронных технологий, достижения в области полупроводниковых гетероструктур и связанной с ними революции в области СВЧ-электроники привели к тому, что частотный диапазон свыше 10 ГГц может стать массовым, т. е. аппаратура для работы в нем становится малогабаритной, воспроизводимой и (при больших объемах выпуска) дешевой.

Системы беспроводного широкополосного доступа своим массовым развитием изначально обязаны телевидению, которое требовало все новых и более высокоскоростных средств доставки ТВ-трафика до зрителей. Кабельные системы со своими задачами справлялись не везде, и в 1970 году Федеральная комиссия связи (FCC) США представила, по-видимому, первую беспроводную многоточечную систему распределения MDS (Multipoint Distribution System). Она предусматривала вещание в радиусе порядка 30 миль от передатчика в диапазоне 2,1–2,7 ГГц. Эксперимент не удался — стоимость оборудования оказалась неконкурентоспособной с кабельным и спутниковым ТВ.

Через четверть века на смену MDS пришли новые системы MMDS (Multichannel (или Microwave) Multipoint Distribution Service) — многоканальные (микроволновые) многоточечные распределительные системы. Они позволяют работать в диапазоне 2,5–2,7 ГГц в зоне радиусом до 35 миль (50–60 км — фактически зона прямой видимости). При этом мощность передатчика базовой станции, как правило, не превышает 100 Вт (до 1 кВт) — сравните с передатчиками 10–50 кВт на ретрансляторах обычных аналоговых телевизионных сетей. В оборудовании MMDS используется квадратурная амплитудная модуляция — от четырехпозиционной QPSK (2 бита на модуляционный символ) до 256-позиционной QAM (256-QAM, 8 бит на символ). Поэтому скорости передачи данных в MMDS могут достигать очень больших значений при соответствующей ширине канала. Так, в стандартном телевизионном канале 8 МГц возможна скорость до 64 Мбит/с, но лишь теоретически. Отметим, что концепция MMDS разрабатывалась для трансляции телевизионных программ, поэтому она принципиально однонаправленная. В последние годы стали появляться сообщения об оборудовании для двунаправленной работы — производители стали добавлять обратный канал, как правило, с гораздо меньшей (в 4–8 раз) полосой пропускания, что позволяет использовать системы MMDS и для обмена данными (доступ в Интернет, интерактивное ТВ, мосты между локальными сетями и т. д.).

Однако сети MMDS так и не стали массовыми — к 2000 году в США они насчитывали около 20 тыс. абонентов [228]. Причина очевидна: оборудование слишком дорогое по сравнению с альтернативными решениями (кабельное телевидение, xDSL-модемы и т. д.). Следующим шагом в области систем широкополосного доступа стало появление концепции Local Multipoint Distribution Service (LMDS) — локальная многоточечная распределительная система. Ее испытания прошли в 1992 году в Нью-Йорке. В 1998 году FCC объявила о начале лицензирования LMDS. Первой реально действующей LMDS-системой стала сотовая телевизионная сеть компании Cellular Vision в районе Брайтон-Бич (этот район Бруклина не был охвачен сетью кабельного телевидения).

Изначально LMDS предназначалась для работы в диапазоне 27,5–29,5 ГГц, однако рабочий диапазон зависит от распределения частот в конкретном регионе.

В Европе появилась аналогичная система MVDS (Multipoint Video Distribution Systems), ориентированная на диапазон 40,5–42,5 ГГц (диапазон, выделенный в Европе для аналогового ТВ-вещания).

Системы LMDS/MVDS называют сотовым телевидением, поскольку радиус действия каждого ретранслятора невелик — порядка 3–8 км. В системах используют относительно маломощные передатчики — не более десятков ватт в групповых передатчиках (до 100–300 мВт на канал). Кроме того, в миллиметровом диапазоне затухание радиоволн весьма велико. Но, с другой стороны, волны этого диапазона отражаются от препятствий с малыми потерями, что можно эффективно использовать в условиях городской застройки, работая на переотраженных сигналах.

Системы LMDS/MVDS используют те же методы модуляции (QPSK, QAM) и частотные планы (19,5–39 МГц), что и системы спутникового ТВ-вещания. Диапазон их работы определяется наличием свободного частотного ресурса и может быть, например, 10, 24, 31, 38 ГГц. При общей ширине полосы 2 ГГц эти системы позволяют передавать от 96 до 128 аналоговых ТВ-каналов. Сотовая структура сетей LMDS/MVDS открывает широкие возможности для частотного планирования, включая такие механизмы, как различная поляризация сигналов, применение направленных (секторных) антенн, использование одних и тех же каналов в разных сотах и т. д. Важно отметить, что в современных трансляционных системах твердотельные выходные усилители усиливают сигнал только в активных каналах, а не во всей полосе сразу. Именно это и обуславливает относительно невысокую мощность излучения групповых передатчиков. Для сравнения — в Нью-Йорке при внедрении системы LMDS в районе Брайтон-Бич использовали выходные усилители на основе ламп бегущей волны мощностью свыше 100 Вт, усиливающие сразу весь рабочий диапазон. Сейчас там 17 базовых станций обеспечивают 48 аналоговых ТВ-каналов (NTSC) [230].

С 2000 года системы LMDS стали двунаправленными, что открыло перед ними широчайшие возможности для миграции из области ТВ-вещания в зону интерактивных приложений, важнейшее из которых — Интернет, а также интерактивное телевидение (Video on Demand). Операторы соответствующих сервисов не замедлили этим воспользоваться. Однако широкому внедрению систем широкополосного доступа мешало отсутствие единого стандарта — аппаратура различных производителей оказывалась несовместимой, специализированная элементная база не могла стать массовой, соответственно, цены оставались высокими. Названия MMDS и LMDS/MVDS фактически обозначали только тип сервиса и самые основные функциональные возможности, а не методы практической реализации. Единой технической концепции (методы модуляции, механизмы доступа к каналам и т. п.) не было. Развитие этих систем резко остановилось, многие известные производители (например, корпорации Nortel, ADC) вообще объявили о прекращении выпуска соответствующего оборудования. Решением возникшей проблемы стал стандарт широкополосных региональных сетей IEEE 802.16.

2.3.1 Региональные сети широкополосного доступа под управлением IEEE 802.16

В августе 1998 года по инициативе Национальной испытательной лаборатории беспроводных электронных систем Национального института стандартов и тех-

нологии США (National Wireless Electronics Systems Testbed of the U.S. National Institute of Standards and Technology) комитет 802 IEEE организовал рабочую группу 802.16. С июля 1999 года группа приступила к регулярной работе над новым стандартом широкополосных городских (региональных) сетей передачи данных (MAN — Metropolitan Access Network) с фиксированным доступом. Соответственно, новый стандарт получил название WirelessMAN. Изначально деятельность велась в трех направлениях — разработка стандартов для диапазона 10–66 ГГц (первоначально обозначался 802.16.1) и 2–11 ГГц (802.16.3), а также стандарта, регламентирующего совместную работу различных систем широкополосного беспроводного вещания (802.16.2).

Уже в декабре 2001 года стандарт IEEE 802.16 «Air Interface for Fixed Broadband Wireless Access Systems» (воздушный интерфейс для фиксированных систем с широкополосным беспроводным доступом) был утвержден (официально опубликован 8 апреля 2002 года). Он описывал общие принципы построения систем широкополосного беспроводного доступа и сосредоточивался на диапазоне 10–66 ГГц. Отметим, что 15 января 2003 года был опубликован документ IEEE 802.16c — поправки и дополнения к IEEE 802.16, касающиеся работы в диапазоне 10–66 ГГц. 10 сентября 2001 года увидел свет стандарт IEEE 802.16.2 «Coexistence of Fixed Broadband Wireless Access Systems» (сосуществование фиксированных систем широкополосного беспроводного доступа). Над более низкочастотным диапазоном работы продолжались чуть дольше — стандарт IEEE 802.16a «Medium Access Control Modifications and Additional Physical Layer Specifications for 2–11 GHz» (модификации управления доступа к среде передачи и дополнительные спецификации физического уровня для диапазона 2–11 ГГц), регламентирующий работу в диапазоне 2–11 ГГц, был утвержден 29 января 2003 года, а 1 апреля опубликован.

Все стандарты группы 802.16 (равно как и их собратья 802.11) описывают два нижних уровня модели взаимодействия открытых систем (OSI) — физический и уровень контроля доступа к среде передачи (MAC — Medium Access Control). В стандартах этой группы идет речь о радиointерфейсах, методах модуляции и доступа к каналам, системе управления потоками, о структурах передаваемых данных, о механизмах связи протоколов передачи данных верхних уровней (прежде всего ATM и IP) с протоколами физического уровня IEEE 802.16 и др.

Разумеется, работать с тремя документами вместо одного неудобно. Кроме того, сразу же после публикации этих документов стали появляться многочисленные исправления и дополнения, только что выявленные на практике. Труд учитывать все эти поправки взяла на себя Рабочая группа 802.16d IEEE. Непосредственно к работе по созданию единого документа с учетом всех поправок она приступила 11 сентября 2003 года (рабочее название draft-версии того периода — 802.16REVd). Почти через год, 24 июня 2004 года был официально утвержден новый стандарт — 802.16-2004, заменяющий собой документы 802.16-2001, 802.16c-2002 и 802.16a-2003. Принципы организации сетей, предусмотренные стандартом IEEE 802.16, мы подробно рассмотрим в главе 8.

2.3.2 Мобильные сотовые технологии

Сотовая связь — одно из революционных достижений в области беспроводных сетей, ставшее обыденным за последние 10 лет. Роль этой технологии в 90-е годы

столь же велика, как бум персональных компьютеров в 80-х. Мобильный телефон превратился в привычный предмет обихода, по стоимости приближающийся к обычному телефонному аппарату (а по распространенности уже превзошедший число телефонных аппаратов фиксированной связи). Широкие возможности сетей мобильной связи не могут не привлекать внимание разработчиков различного рода систем — мониторинговых, охранных, коммуникационных и т. д. Смена же поколений сотовой связи столь стремительна, что достижимые технические возможности ощутимо опережают реальную потребность пользователей, это ярко демонстрирует проблема с внедрением сотовых сетей третьего поколения (3G). А специалисты говорят уже о 4G.

Рост сотовых сетей связи носит поистине «взрывной» характер. Число их абонентов уже превысило миллиард — примерно столько, сколько всего абонентов проводной связи. В 2001 году насчитывалось около 430 млн абонентов сетей GSM (69% всех абонентов мира), 65 млн приверженцев CDMA, 47,1 млн пользователей DAMPS, еще около 58 млн абонентов сетей других стандартов [228]. В России на 1 января 2001 года было порядка 3,4 млн абонентов — годовой рост составил тогда 152% [230]! К марту 2004 года их число выросло более чем в 10 раз, достигнув 42,37 млн абонентов.

Очевидно, что в подобной ситуации потребность в едином стандарте или, по крайней мере, в совместимых стандартах весьма остра. К тому же сотовая связь все активнее вторгается в сферу передачи данных — это и электронная почта, и доступ к ресурсам Интернета, и обмен видеоинформацией, и т. п. Но, чтобы подобные услуги оказались востребованными, необходимы иные скорости обмена (т. е. превосходящие стандартные для сетей второго поколения (2G) 9,6 или 14,4 кбит/с) и новые принципы организации соединений.

Еще в 1992 году Международный союз электросвязи (ITU) инициировал работы над стандартом «всемирной подвижной радиосвязи» IMT-2000. Предполагалось, что к 2000 году появится спецификация сетей подвижной связи, действующая в диапазоне около 2000 МГц и со скоростью передачи данных порядка 2 Мбит/с. Одним из требований была возможность точного определения местоположения мобильных терминалов. Причем речь шла о единой системе телекоммуникаций, объединяющей спутниковые, мобильные, фиксированные виды связи. Надеждам этим сбыться не было суждено, однако некая определенность со стандартами сетей третьего поколения наступила. Но, прежде чем говорить о 3G, рассмотрим существующие стандарты сотовой связи.

Глобальная система мобильной связи (GSM)

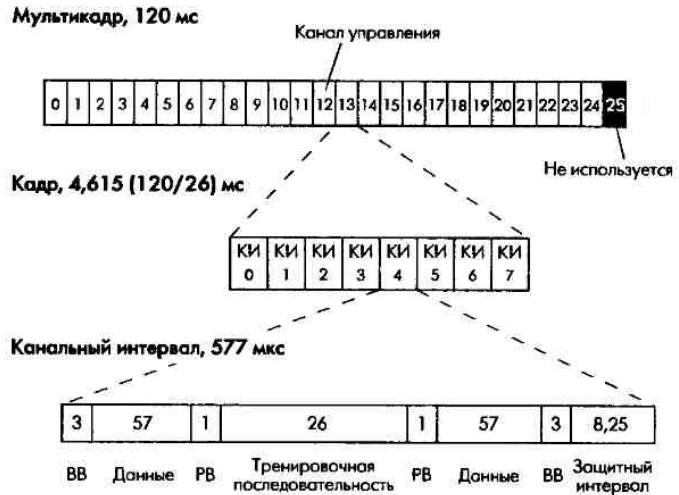
Безусловный лидер по распространенности на мировом рынке — стандарт GSM. Его история началась в 1982 году, когда Европейская конференция администраций почты и телеграфа (CEPT) создала рабочую группу GSM (Group Special Mobile) для разработки общеевропейской системы подвижной сотовой связи. В 1989 году работы по GSM перешли под эгиду Европейского института стандартизации электросвязи (ETSI), и в 1990 году были опубликованы спецификации первой фазы стандарта. В 1993 году в 22 странах мира действовало 36 сетей GSM. К 1995 году насчитывалось около 5 млн абонентов, стандарт стал общемировым и расшифровывался уже как Global System for Mobile Communications. За последую-

шие шесть лет число абонентов возросло в 84 раза, что составляет порядка 70% пользователей сотовой связи во всем мире. Примечательно, что такая же доля сторонников GSM и в России.

GSM действует в диапазонах 900 и 1800 МГц (в США — 1900 МГц). В Европе и России в диапазоне 900 МГц мобильный телефон передает (восходящий канал) в полосе 890–915 МГц, принимает (нисходящий канал) в интервале 935–960 МГц (для GSM-1800 — 1710–1785 и 1805–1880 МГц соответственно). Весь диапазон делится на частотные каналы по 200 кГц — в GSM-900 всего 124 канала (124 восходящих и 124 нисходящих), разнос между восходящим и нисходящим каналом — 45/95 МГц (в диапазонах 900/1800 МГц, соответственно). Базовая станция поддерживает от 1 до 16 частотных каналов. Таким образом, в GSM реализован частотный метод дуплексирования каналов (FDD).

Что касается доступа к среде передачи, в GSM использован принцип временного разделения канала — TDMA. Частотные каналы разбиты на кадры по 8 временных интервалов (канальные интервалы) длительностью по 577 мкс. Каждому физическому каналу соответствует один определенный временной интервал на определенной частоте. Таким образом, мобильный терминал (МТ) передает базовой станции (БС) информацию в течение 577 мкс каждые 4615 мкс. БС связывается с МТ точно так же, но на три временных интервала раньше МТ (и на частоте на 45 МГц выше), чтобы разнести во времени прием и передачу. Это существенно упрощает аппаратуру МТ.

Рис. 2.40. Временное разделение каналов в GSM



Временные интервалы в GSM бывают пяти типов — нормальный, подстройки частоты, синхронизации, установочный и доступа. Структура нормального временного интервала показана на рис. 2.40. Полезная информация передается двумя блоками по 57 бит. Между ними расположена тренировочная последовательность в 26 бит, ограниченная одноразрядными указателями PB (Pointer Bit). Интервалы ВВ (Border Bit) длиной 3 бита ограничивают всю передаваемую последовательность. После трансляции всех 148 бит канального интервала передат-

чик «молчит» в течение защитного интервала ST (Shield Time) длительностью 30,44 мкс, что по времени эквивалентно передаче 8,25 бит.

Каждые 26 кадров объединены в мультикадр продолжительностью 120 мс. В мультикадре каждый 13-й кадр зарезервирован для канала управления, а в течение каждого 26-го кадра вся система «молчит».

Отметим, что в GSM использован принцип медленных частотных скачков — прием/передача нового кадра может происходить на новой несущей частоте. При этом сохраняется дуплексный разнос в 45 МГц. Начальное значение несущей и последовательность изменения назначаются мобильному терминалу при установлении связи. Модуляция сигнала — двоичная гауссова с минимальным частотным сдвигом GMSK (один бит на символ).

Радиус соты в GSM — до 35 км — ограничен возрастающей временной задержкой распространения сигнала, к которой чувствительна технология TDMA. Сетевая инфраструктура GSM/MAP основана на системе сигнализации OKC7 (SS7) [30]. Для кодирования речи применен кодек VCELP на основе алгоритма RPE-LTP (Regular Pulse Excitation-Long Term Prediction) со скоростью 13 кбит/с. Скорость передачи данных — до 9,6 кбит/с (по стандартной схеме).

Стандарт AMPS/DAMPS

13 октября 1983 года в Чикаго заработала первая сеть стандарта AMPS (Advanced Mobile Phone System). Родившись на Американском континенте, этот стандарт распространился по всему миру, попав и в Россию. AMPS рассчитан на диапазон 824–840 и 869–894 МГц, каналы — дуплексные с разносом на 45 МГц, ширина канала 30 кГц. Как стандарт первого поколения, AMPS потребовал модернизации. Так, в 1988 возник DAMPS — Digital AMPS, цифровой стандарт второго поколения, действующий в том же диапазоне, что и предшественник. Ширина канала в DAMPS — те же 30 кГц, но применено временное разделение каналов — циклически повторяющиеся кадры с тремя временными интервалами. Речевой кодек — VCELP, 8 кбит/с. Поскольку стандарт американский, его сетевая инфраструктура — ANSI-41. Размер соты — до 20 км. В отличие от аналогового прародителя, DAMPS особого распространения в мире получить не успел и уже никогда не успеет.

Стандарт CDMA

CDMA расширяется как множественный доступ с кодовым разделением каналов (Code-Division Multiple Access). Сама по себе технология не нова: первая в СССР работа на эту тему — «Основы теории линейной селекции» Д.В. Агеева — была опубликована в сборнике ЛЭИС в 1935 году. Значительно продвинули технологию работы К. Шеннона. До определенного момента CDMA находил применение только в военной и специальной технике из-за сложности аппаратуры для обработки сигналов. Зато такие свойства технологии, как высокая стойкость к помехам и скрытность передачи, в данной области оказались незаменимыми.

С развитием микроэлектроники стало возможным создание недорогих портативных станций CDMA. Лидер в этой области — американская компания Qualcomm, разработавшая спецификацию IS-95 (cdmaOne). Возможно, именно IS-95

и уготована самая долгая жизнь, поскольку на базе этого стандарта развивается одно из направлений сотовой телефонии третьего поколения.

Упрощенно рассмотрим принцип действия CDMA. Различают три вида кодового разделения каналов — расширение спектра методом прямой последовательности (DS), частотных скачков (FH) и временных скачков (TH) [68]. Нас интересует метод DS, в отечественной литературе его называют передачей на основе шумоподобных сигналов (ШПС). В CDMA-DS каждый бит информационного сигнала заменяется некоторой фиксированной последовательностью определенной длины — базой сигнала. Ноль и единица могут, например, кодироваться инверсными последовательностями. Для каждого канала задается определенная последовательность (код). Спектр сигнала расширяется пропорционально длине базы. Последовательности обычно подбирают ортогональными (скалярное произведение равно нулю). В приемнике происходит вычисление корреляционных интегралов входного сигнала и кодовой последовательности определенного канала. В результате принимается только тот сигнал, который был расширен посредством заданной кодовой последовательности (корреляционная функция выше порогового значения). Все остальные сигналы воспринимаются как шум. Таким образом, в одной полосе могут работать несколько приемопередатчиков, не мешая друг другу. Благодаря широкополосности сигнала снижается его мощность, причем при очень длинной базе — ниже уровня белого шума.

Сильно возрастает помехоустойчивость, а с ней и качество связи — узкополосная помеха не повлияет на широкополосный сигнал. Кодовая последовательность автоматически является и элементом криптозащиты. Что особенно привлекательно для операторов сотовой связи — упрощается проблема частотного планирования, поскольку все станции работают в одной полосе. Все эти свойства и предопределили успех CDMA.

Естественно, принцип взаимодействия базовой и мобильной станций в стандарте IS-95 гораздо сложнее. Рассмотрим его немного подробнее, поскольку именно этот стандарт лег в основу ряда сетей третьего поколения.

Сети IS-95 занимают практически тот же частотный диапазон, что и сети AMPS: 824–840 и 869–894 МГц. Нисходящий канал (от БС к МТ) всегда на 45 МГц выше восходящего. Ширина канала — 1,25 МГц. Существует и более высокочастотная версия в диапазонах 1890–1930 и 1950–1990 МГц. Там дуплексный разнос — 80 МГц. Ниже мы рассмотрим работу в диапазоне до 900 МГц — в более высокочастотной версии все аналогично, только скорость передачи данных в 1,5 раза выше: до 14,4 кбит/с.

Нисходящий канал содержит 64 логических канала. Логические каналы формируются за счет расширения спектра сигнала последовательностями Уолша (Walsh). Каждая из этих последовательностей представляет собой одну из 64 строк матрицы Адамара (Hadamar). Основное их свойство в том, что все строки матрицы (и их инверсия) взаимно ортогональны.

Способ построения матрицы Адамара прост. Матрица первого порядка $A_1 = [1]$. Матрица A_{2n} образуется по схеме

$$A_{2n} = \begin{bmatrix} A_n & A_n \\ A_n & -A_n \end{bmatrix}.$$

Так, матрица Адамара второго порядка имеет вид

$$A_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix};$$

матрица Адамара четвертого порядка

$$A_4 = \begin{bmatrix} A_2 & A_2 \\ A_2 & -A_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

В стандарте IS-95 используются матрицы Адамара 64-го порядка. Последовательность Уолша отличается от строки матрицы Адамара только тем, что в ней -1 заменена на 0 .

Рассмотрим процесс передачи в нисходящем канале (рис. 2.41). Входной поток (данные, оцифрованный голос) (1,2–9,6 кбит/с) подвергается защитному сверточному кодированию с скоростью $1/2$ и попадают в повторитель, который в зависимости от условий связи может повторять передачу одного блока данных до восьми раз. Затем данные поступают в блок перемежения, защищающий от групповых ошибок. Фактически, это матрица, которую информационные биты заполняют по строкам, а выводятся по столбцам.



Рис. 2.41. Схема передачи в прямом канале *cdmaOne*

Далее поток перемножается с 42-разрядным числом, так называемой маской длинного кода, фактически — идентификационным номером мобильной станции (речь идет о канале передачи трафика, в каналах другого типа маска может формироваться иначе). Это элемент дополнительной криптозащиты. Наконец, поток расширяется посредством последовательностей Уолша (каждый бит перемножается на 64-разрядную последовательность). Каждому из 64 каналов соответствует определенная последовательность. Первая последовательность Уолша закреплена за пилотным каналом.

После расширения последовательностями Уолша скорость потока становится 1,2288 Мбит/с (если быть точным — не бит, а чипов, поскольку бит — понятие информационное, а элементы модулированных последовательностей называют чипами). В результате каждому информационному биту исходного потока соответствует 128 чипов выходной последовательности. Выигрыш в отношении сигнал/шум для расширенного и исходного сигнала составляет $10 \lg 128 = 21$ дБ. Если принять, что на входе приемника допустимо соотношение сигнал/шум в

3 дБ, то передачу теоретически можно вести при уровне сигнала на 18 дБ ниже уровня интерференционных помех.

Прежде чем попасть на модулятор, сигнал дополнительно расширяется псевдослучайной последовательностью с так называемым коротким кодом (период — $2^{15} - 1$) и раскладывается на квадратурные составляющие. Несущая модулируется методом четырехпозиционной фазовой манипуляции — QPSK. Поскольку последовательности Уолла взаимно ортогональны, интерференционные помехи между каналами одной БС практически отсутствуют. Передача ведется пакетами длительностью 20 мс.

Восходящий канал делится на $2^{12} - 1$ логических каналов. Каждой мобильной станции присвоен свой уникальный логический канал на основе 42-битного идентификационного номера. Спектр сигнала в обратном канале расширяется на основе так называемых m -последовательностей (длинного кода). Их генерация происходит в 42-разрядном сдвиговом регистре с обратными связями (рис. 2.42).

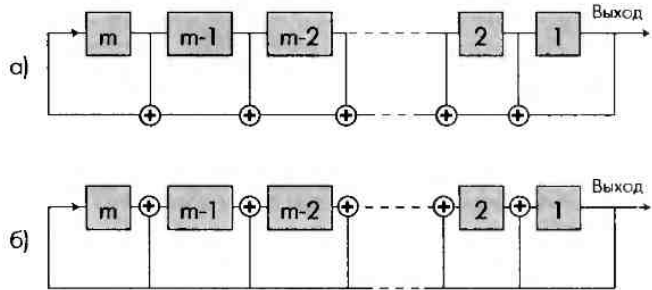


Рис. 2.42. Пикосеть с одним подчиненным устройством (а), несколькими (б) и распределенная сеть (в)

Если начальные значения во всех m узлах регистра не равны 0, генератор на его основе будет выдавать периодическую псевдослучайную последовательность длиной $2^m - 1$. Отличительная особенность m -последовательностей — сумма по модулю 2 (операция «исключающее ИЛИ») m -последовательности с той же последовательностью, смещенной по фазе, дает ту же самую последовательность, но с другим фазовым сдвигом ($M(t/T) \oplus M(t/T + \varphi_1) = M(t/T + \varphi_2)$). Благодаря этому свойству, с помощью идентификационного номера — маски длинного кода — можно задавать начальную фазу последовательности (рис. 2.43). Таким образом, вся система использует один вид псевдослучайной последовательности (ПСП) с очень большим периодом повторения, а селекция логических каналов происходит за счет выбора ее фазы. ПСП обладают всеми свойствами случайных последовательностей, с высокой автокорреляцией при совпадении фаз.

Передача в восходящем канале во многом аналогична передаче в нисходящем (рис. 2.44). Входная информация после сверточного кодирования (со скоростью кодирования 1/3), повторителя и блока перемежения попадает в блок ортогональной модуляции,

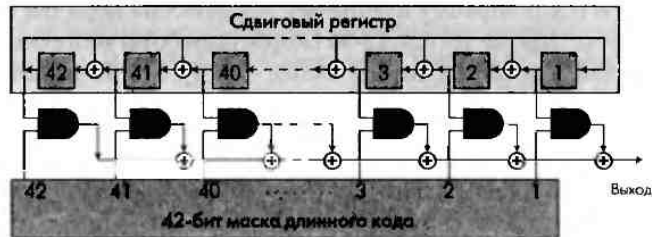


Рис. 2.43. Генератор длинного кода с заданным фазовым сдвигом в cdmaOne

где каждая группа из 6 бит заменяется соответствующей 64-разрядной последовательностью Уолша. Далее поток последовательностей Уолша перемножается на ПСП (длинный код). При этом каждый элемент последовательности Уолша преобразуется в четыре элемента ПСП. Затем поток квадратируется посредством так называемого короткого кода с периодом $2^{15} - 1$. Короткий код необходим для первичной синхронизации МТ с БС.



Рис. 2.44. Схема передачи в обратном канале cdmaOne

Получающиеся в итоге две последовательности имеют период повторения порядка 2^{57} , что при скорости цифрового потока 1,2288 Мчип/с эквивалентно 3700 годам. Они используются для модуляции несущей. В восходящем канале применяется квадратурная фазовая модуляция со сдвигом О-QPSK (см. рис. 2.44), каждому символу соответствуют два бита. Этот вид модуляции позволяет снизить требования к линейности усилителей передающего тракта МТ. В результате всех преобразований каждый бит исходного сообщения заменяется 256 элементами транслируемой последовательности.

Прием сигналов происходит в обратном порядке. Для выделения «своего» сигнала используют цифровые корреляторы, вычисляющие корреляционную функцию с заданной последовательностью Уолша (в МТ) либо с m -последовательностью (БС) в заданной фазе. МТ обычно содержит несколько корреляторов для работы одновременно с несколькими базовыми станциями. Это важно при переходе из соты в соту, когда терминал принимает сигналы от различных БС и, сравнивая их качество, выбирает предпочтительную. Кроме того, несколько корреляторов обеспечивают прием при многолучевом распространении сигнала, что может улучшить качество связи.

Важная особенность стандарта IS-95 — гибкое управление мощностью излучения МТ. В пределах соты уровни принимаемых БС сигналов должны быть одинаковыми независимо от удаления МТ. Для этого мощность МТ регулируется по специальному алгоритму в диапазоне порядка 80 дБ с шагом 1 дБ каждые 1,25 мс. Кроме того, в IS-95 скорость работы голосового кодека не постоянна, как в GSM, а может меняться в зависимости от интенсивности речи от 8 до 1,2 кбит/с. Эти особенности позволяют очень гибко регулировать загрузку в сети, не загружая соту избыточной информацией.

Одна БС может поддерживать до 64 каналов. Однако часть из них — служебные: пилотный, синхронизации, вызова. Оказывают влияние и соседние БС. Однако при фиксированной связи БС поддерживает до 40–45, при подвижной — до 25 каналов передачи трафика. И все это на одной частоте! Технология CDMA требует точной, до микросекунд, синхронизации БС. Для этого используют сиг-

налы глобальной системы позиционирования GPS. Радиус соты — до 20 км, сетевая инфраструктура — ANSI-41.

Третье поколение сотовой связи

Основной недостаток систем мобильной связи второго поколения — низкая скорость передачи данных — 9,6–14,4 кбит/с. В рамках же IMT-2000 стояла задача достичь в сетях 3G скорости потока до 2 Мбит/с для малоподвижных абонентов и до 384 кбит/с — для мобильных. В мире сформировались два глобальных партнерских объединения, формирующих стандарты 3G, — 3GPP и 3GPP2 (3G Partnership Project). В первое вошли ETSI (Европа), подкомитет P1 телекоммуникационного комитета ANSI (США), ARIB и TTC (Япония), SWTS (Китай) и TTA (Южная Корея). Участники 3GPP сумели согласовать особенности своих подходов к технологии широкополосной CDMA (WCDMA) с частотным (FDD) и временным (TDD) дуплексированием, представив ITU проекты IMT-DS и IMT-TC соответственно. В основу легло европейское предложение UTRA (UMTS Terrestrial Radio Access, радиointерфейс наземного доступа к системе UMTS) — UTRA FDD и UTRA TDD. Отметим, что в качестве одного из стандартов IMT-2000 предложено дальнейшее развитие технологии микросотовых сетей DECT (проект IMT-FT).

Члены объединения 3GPP2 предлагают фактически эволюционный путь — варианты развития технологий DAMPS (UWC-136) и cdmaOne (cdma2000). Данные предложения представлены ITU как проекты IMT-SC и IMT-MC.

Таким образом, намечилось два пути: революционный — там, где есть свободный частотный ресурс, и эволюционный — в остальных регионах. Рассмотрим их.

В 1996 году в городе Чиста (Швеция) компания Ericsson запустила первую опытную сеть с технологией WCDMA. Эта технология легла в основу проекта наземного мобильного сегмента европейской универсальной системы телекоммуникаций UMTS. Было предложено два варианта WCDMA — с частотным и временным разносом прямого и обратного каналов (FDD WCDMA и TDD WCDMA) соответственно для парного (предполагается 2110–2170 и 1920–1980 МГц) и непарного спектра частот.

Технология основывается на расширении спектра методом прямой последовательности в полосе 5 МГц на канал. Изначально определенная скорость потока чипов 4,096 Мчип/с для согласования с другими стандартами была снижена до 3,84 Мчип/с. Таким образом, система может поддерживать требуемые 2 Мбит/с для малоподвижных абонентов и 384 кбит/с — для мобильных. Предусмотрена возможность применения интеллектуальных антенных систем (Smart-антенн с цифровым формированием диаграммы направленности). Принципы технологии FDD WCDMA во многом аналогичны cdmaOne (конечно, WCDMA — гораздо сложнее). Одно из принципиальных отличий — сеть на базе FDD WCDMA может быть асинхронной (возможен и синхронный режим).

Для случаев, когда спектральный диапазон ограничен — нет возможности выделять частоты под парные каналы 5 МГц, — проработана версия WCDMA TDD с временным дуплексированием каналов. Принцип прост: весь временной диапазон представляет последовательность равных канальных интервалов. В течение каждого из них в каждом из логических каналов (с кодовым разделением)

происходит передача в одном направлении – от БС или от МТ. Таким образом, в определенные промежутки все каналы – либо восходящие, либо нисходящие. Соотношение и последовательность восходящих/нисходящих канальных интервалов может гибко изменяться в зависимости от интенсивности трафика в обе стороны. Это крайне важно для многих приложений с асимметричной передачей данных (например, доступ в Интернет). По сравнению с FDD WCDMA сети с TDD должны быть синхронными, в остальном же их параметры практически совпадают.

Развитием метода WCDMA TDD стала система TD-SCDMA, созданная совместно компанией Siemens и китайской Академией телекоммуникационных технологий (China Academy of Telecommunications Technology – CATT). Это стандарт физического уровня беспроводных сетей 3G, одобренный ITU и объединением стандартизирующих организаций 3GPP как часть пула стандартов UMTS. TD-SCDMA (технология CDMA с одной несущей и временным дуплексированием) ориентирована для работы в зонах с высоким дефицитом частотного ресурса – именно такова ситуация в КНР, связанная с высочайшей плотностью населения (в несколько раз выше, чем в густонаселенной Европе).

Сама технология доступа представляет собой комбинацию трех механизмов: временного разделения дуплексных каналов (TDD), временного мультиплексирования каналов (TDMA) и кодового мультиплексирования каналов (CDMA). Обмен происходит циклически повторяющимися кадрами (фреймами) длительностью 5 мс, разделенными на семь временных интервалов (тайм-слотов). Кроме того, в каждом тайм-слоте возможно формирование до 16 CDMA-каналов на основе 16 кодовых последовательностей. Важнейшая особенность – предусмотрена возможность гибкого распределения тайм-слотов исходя из фактически передаваемого трафика. Например, в асимметричных приложениях (доступ в Интернет) для восходящего канала можно выделить один тайм-слот, для нисходящего – остальные шесть.

Ширина одной полосы TD-SCDMA 1,6 МГц. Скорость передачи модуляционных символов 1,28 Мчип/с. Это, вместе с переменным числом тайм-слотов во фрейме, назначенных одному соединению, позволяет добиваться скорости передачи данных в широчайшем диапазоне: от 1,2 кбит/с до 2 Мбит/с. Заявленная дальность передачи – 40 км, допустимая максимальная скорость движения мобильного абонента – не менее 120 км/ч.

Важнейшее достоинство TD-SCDMA – эффективное использование спектра. В технологиях с частотным разносом восходящего/нисходящего каналов на одно соединение всегда выделяются две частотные полосы. И проблема не только в том, что эти две полосы с разносом в 45–220 МГц еще надо найти. При асимметричной передаче (а именно таковы многие мультимедийные приложения) частотный ресурс одного из каналов в большой степени фактически пропадает, поскольку для соединения назначается два частотно-разнесенных канала: приемный и передающий. В случае TD-SCDMA такого не происходит, поскольку частотная полоса одна и соотношение входящего/нисходящего трафика можно гибко варьировать.

Не менее важно, что разработчики TD-SCDMA предусмотрели ее гибкую интеграцию с GSM-сетями, а также мягкий переход к WCDMA-сетям благо-

даря поддержке сигнализации и протоколов верхних уровней как GSM, так и WCDMA. Более того, первые телефоны стандарта TD-SCDMA были двухмодовыми, на основе GSM-чипсета с дополнительной СБИС поддержки TD-SCDMA.

WCDMA (UMTS) изначально разрабатывалась как замена сетей GSM с возможностью плавного перехода. Поэтому ее сетевая инфраструктура совместима с MAP/GSM. Кроме того, она ориентирована на глобальные сети с пакетной коммутацией (IP, X.25). Операторы могут создавать «островки» WCDMA в особо густонаселенных районах, постепенно расширяя их. Поэтому все абонентские терминалы для WCDMA в Европе будут поддерживать GSM. Однако первая сеть WCDMA начала действовать в Японии (оператор — компания NTT DoCoMo) в 2002 году, где телефонов GSM никогда не было. Для японских операторов WCDMA привлекательна из-за ее высокой абонентской емкости.

Уже действующие сети — это построенная инфраструктура, сформированная и развивающаяся абонентская база и, что самое главное, — огромные вложенные средства и определенная инвестиционная перспективность (если, конечно, сеть успешна). Не менее важно, что за оператором сети закреплён частотный диапазон. Поиск путей усовершенствования происходит постоянно. Однако, прежде чем строить принципиально новую сеть (WCDMA), необходимо рассмотреть возможности развития существующих.

Значительный потенциал заложен в стандарте IS-95. Прямым его развитием стала спецификация IS-95b. Она позволяет объединять до восьми логических каналов. Теоретически достижимая скорость при этом $14,4 \times 8 = 115,2$ кбит/с. Реально работающие сети IS-95b обеспечивают передачу до 64 кбит/с.

Следующий шаг развития IS-95 — проект cdma2000, который в итоге должен удовлетворять требованиям IMT-2000. Предусматривалось три стадии развития cdma2000: 1X, 3X и cdma2000 DS (прямая последовательность). Последний вариант технически аналогичен WCDMA, и потому работы над ним были прекращены.

CDMA 1X (CDMA 1XRTT) позволяет увеличить число логических каналов до 128 в той же спектральной полосе 1,25 МГц. При этом реальная скорость — до 144 Мбит/с. Первая такая сеть была организована в Южной Корее (оператор — SK Telecom).

Компания Qualcomm предложила технологию увеличения скорости HDR (High Data Rate). Ее основная идея — гибкое увеличение числа положений вектора сигнала (символов) при фазовой модуляции, следовательно, увеличение числа бит на символ. В стандартной модуляции несущей в CDMA (квадратурная фазовая, QPSK) используются четыре символа, каждый определяет два бита. Если увеличить число возможных символов до восьми (8PSK), каждый отсчет сигнала будет определять три бита и скорость передачи возрастет в полтора раза. Очевидно, что чем хуже условия связи, тем меньше символов может распознать приемник (т.е. тем больший фазовый сдвиг способен распознать его детектор). Поэтому в зависимости от зашумленности эфира вид модуляции в HDR меняется. Технология HDR позволяет достигать пиковых скоростей порядка 2,4 Мбит/с в стандартной полосе 1,25 МГц. О работе над аналогичной технологией — 1Xtreme — заявила и компания Motorola.

Спецификация CDMA 3X — вторая фаза проекта cdma2000. Обозначение 3X указывает на утроение спектральной полосы канала cdmaOne: $1,25 \times 3 =$

= 3,75 МГц. При этом в обратном канале происходит передача методом прямой последовательности в полосе 3,75 МГц. В прямом же канале данные передаются параллельно по трем стандартным IS-95 каналам шириной 1,25 МГц (технология с несколькими несущими, MC) (рис. 2.45). В результате скорость может превышать 2 Мбит/с. Поскольку технология базируется на IS-95, БС в сетях cdma2000 требуют синхронизации. Существенно, что вполне возможно дальнейшее масштабирование: 6X, 9X и т. д. с соответствующим ростом производительности или емкости.

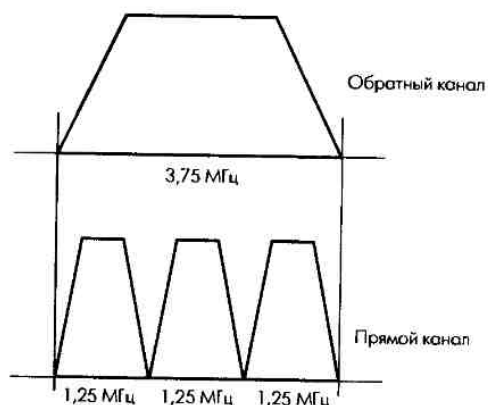


Рис. 2.45. Спектр сигналов в технологии cdma2000 3X

Важнейшая особенность cdma2000 — его полная совместимость с предыдущими фазами, вплоть до cdmaOne. Не требуется нового частотного диапазона и существенного изменения аппаратуры. Фактически возможно плавное улучшение параметров существующих сетей IS-95 до требований IMT-2000. Правда, возникает глобальная проблема: совместимость стандартов cdma2000 и WCDMA. Причем как на уровне терминалов, так и сетевой инфраструктуры (MAP/GSM и ANSI-41). Впрочем, обе технологии поддерживают протокол IP, что может стать для них объединяющей платформой.

Отметим, что, поскольку в России сетей CDMA практически нет, казалось, что вариант cdma2000 с его достоинствами — не для нас. Однако это утверждение опровергнуто появлением сетей стандарта cdma2000-450, работающих в диапазоне 450 МГц. Они пришли на смену сетям аналогового сотового стандарта NMT-450.

Плавный переход к сетям 3G крайне привлекателен и для операторов сетей с технологией TDMA — GSM и DAMPS. Методов повышения быстродействия сетей GSM несколько. Изначальная скорость передачи данных в них составляла 9,6 кбит/с. Однако по каждому каналу через 4,615 мс передается эквивалент 156-разрядного пакета, следовательно, максимально возможная теоретическая скорость в GSM-канале 33,8 кбит/с. Существенная доля этой полосы отведена для служебной информации, сигнализации, а также алгоритмов защиты от ошибок и криптозащиты. Оставшаяся часть полосы используется для передачи оцифрованной речи со скоростью 13 кбит/с. Скорость передачи данных еще ниже, поскольку при стандартной схеме они следуют через речевой кодек. Изменив алгоритм защитного кодирования, удастся увеличить скорость до 14,4 кбит/с.

Следующим шагом стало введение схемы HSCSD (высокоскоростная передача данных по коммутируемым каналам). Она предусматривает объединение нескольких канальных интервалов. Так, при объединении двух интервалов возможна скорость 19,2 ($9,6 \times 2$) и 28,8 ($14,4 \times 2$) кбит/с. Для этого в основном необходимы изменения в программах поддержки протоколов, не затрагивающие аппаратной части и инфраструктуры сети. Более высокие скорости (например,

$9,6 \times 4 = 38,4$ кбит/с) требуют модернизации аппаратуры мобильных телефонов. Дальнейшее увеличение скорости, например до 76,8 кбит/с ($9,6 \times 8$), ограничено сетевой инфраструктурой (64 кбит/с в канале между БС и коммутатором).

«Прорывным» стало внедрение технологии пакетной передачи GRPS — дальнейшего развития HSCSD. В самом деле, для сотовых систем второго поколения сеть с коммутацией каналов — атавизм. Введение пакетной коммутации делает мобильные сети легко совместимыми с IP- и X.25-сетями, создавая тем самым прекрасную платформу для перехода к WCDMA.

При пакетной коммутации данные передаются через свободные от речевого трафика каналные интервалы. Совершенно реальными становятся скорости свыше 100 кбит/с (теоретический предел — $33,8 \times 8 = 270,4$ кбит/с). Мобильный терминал может одновременно поддерживать голосовое соединение и обмен данными без ухудшения качества речи.

Одно из важнейших достоинств пакетной коммутации — очень быстрое установление соединения. Абонент занимает канал только в момент передачи. Поэтому тарификация может происходить на основе реально переданной информации (числа пакетов), а не пропорционально времени нахождения в сети, как при коммутации каналов. Скорости обмена могут гибко меняться. Кроме того, пакетный режим позволяет мобильному телефону работать как персональная радиостанция в режиме постоянно установленного соединения (виртуального), т.е. можно разговаривать, не нажимая ни на какие кнопки. Голос при этом передается по IP-протоколу в пакетном режиме, используя технологию GPRS. Поскольку трафик расходуется только при непосредственной передаче данных, все остальное время телефон остается «на связи» (точнее, в состоянии постоянной готовности к связи, собственно соединение при GPRS устанавливается за пренебрежимо малое время), но пользователь за это не платит.

Чтобы внедрить технологию GRPS в существующие сети GSM, их инфраструктуру достаточно дооснастить оборудованием пакетной передачи (основные устройства — шлюзовые и управляющие узлы, GGSN и SGSN), а каждый GSM-контроллер — блоками управления пакетной связью (PCU). Развивать GRPS-сеть можно постепенно, оснащая ее узлами GGSN и SGSN.

Дальнейшим развитием пакетной передачи стала технология EDGE (Enhanced Data for Global Evolution, изначально вместо Global стояло GSM) [170, 178]. В ее основе — изменение метода модуляции несущей и адаптивная схема защитного кодирования. Напомним, в GSM применяется модуляция GMSK с одним битом на символ.

В EDGE предусмотрена модуляция 8PSK с тремя битами на символ. Таким образом, скорость передачи утраивается. Предусмотрено два режима EDGE: с коммутацией пакетов (EGRPS, Enhanced GRPS) и с коммутацией каналов (ECSD, Enhanced Circuit Switched Data). Скорости в одном канале передачи увеличиваются соответственно до 69,2 и 38,4 кбит/с. В режиме коммутации каналов возможно объединение каналных интервалов, как в технологии HSCSD.

Режим пакетной передачи предусматривает девять скоростей [170], отличающихся схемой защитного кодирования и видом модуляции несущих. Скорость автоматически изменяется от пакета к пакету в зависимости от условий в эфире. На физическом уровне протокол EDGE совпадает с GSM, включая структу-

ру кадров и мультикадра (только разряды РВ включены в блоки данных, см. рис. 2.40). При этом общая скорость на несущую — до 384 кбит/с, что позволяет рассматривать EDGE как технологию сетей третьего поколения.

EDGE может послужить базой для эволюции не только сетей GSM, но и DAMPS. В январе 1998 года консорциум UWCC (Universal Wireless Communications Consortium) принял EDGE за основу дальнейшего развития сетей DAMPS, создав концепцию UWC-136. Она предусматривает расширение полосы канала DAMPS с 30 до 200 кГц с соответствующим ростом скорости. Данный проект внесен на рассмотрение ITU как система 3G (проект IMT-SC). На платформе EDGE возможна и интеграция сетей DAMPS с сетями GSM.

Отметим, что прорабатываются два варианта EDGE-сетей: COMPACT и Classic. Последний ориентирован на стандартные GSM-системы, тогда как COMPACT — это решение в условиях ограниченного частотного ресурса. Сети COMPACT могут занимать диапазон шириной всего в 600 МГц. При этом БС должны быть синхронизированы, например, с помощью сигналов системы GPS.

Естественно возникает вопрос: нужны ли сети 3G? Тревогу вызвали результаты уже первых аукционов по продаже лицензий на право предоставлять услуги 3G. В Великобритании общая сумма продаж лицензий составила 34 млрд долл., в Германии — 46 млрд долл. Возмещать операторам такие затраты будут пользователи их услуг. Захотят ли они ими пользоваться, если есть более дешевые альтернативные решения?

Готовность рынка принять данную технологию не велика. Это подтверждают события во Франции, где на четыре 3G-лицензии нашлось лишь два оператора-претендента при «умеренной» цене лицензии — около 4,5 млрд долл. Кроме того, технические возможности и пользовательские приложения — вещи разные. Просмотр видеороликов в мчащемся автомобиле потрясает воображение инженера, но нужен ли такой сервис массовому пользователю? Тем более на малом экране портативного устройства.

С другой стороны, достижения производителей элементной базы для устройств 3G не могут быть проигнорированы рынком. Кроме того, сети 3G нельзя рассматривать просто как беспроводную телефонию. Это — высокоскоростные БСПИ со всеми вытекающими последствиями. Важнейшее из них — в восемь раз более эффективное, по оценкам экспертов, использование емкости сети [82]. Разумеется, при наличии соответствующих информационных сервисов. Это означает, что падает удельная себестоимость инфраструктуры сетей. Так, если затраты для ввода одного абонентского номера GSM составляют, по разным оценкам, 320–350 долл., то при эволюционном переходе к сетям UMTS дополнительные затраты не превышают 50 долл., а потребности в услугах передачи данных неуклонно растут. Так, в августе 2003 года средний доход от передачи данных у операторов Западной Европы составил 15% от общего объема и продолжает расти. Предполагается, что в 2005 году, с учетом распространения услуг 3G-сетей, этот показатель составит 30%.

Сегодня лидерство во внедрении 3G уверенно захватили страны Юго-Востока. 2003 год можно назвать первым годом эксплуатации 3G. В 2002 году японская корпорация NTT DoCoMo первой построила коммерческую 3G-сеть и начала активно оказывать услуги. Через год ее 3G-услугами пользовались 6,8 млн чело-

век. Однако этого крупнейшего японского оператора обогнала компания KDDI, стартовавшая позднее. За полтора года к ее 3G-сети подключилось более 10 млн абонентов. Заметим, что к началу 2003 года общее число абонентов 3G в Японии составляло лишь 7,161 млн, причем всего абонентов мобильной связи в этой стране тогда было 75,656 млн человек. В конце же 2003 года услугами связи 3G пользовались более 60 млн человек, и ежемесячно их число увеличивается более чем на 3,5 млн. Всего коммерческие услуги 3G предоставляются более чем в 30 странах мира: развернуто 70 сетей стандарта cdma2000 и 12 сетей стандарта UMTS. В 2003 году доля мирового рынка производимого оборудования 3G составляла 30%, к 2005 году предполагалось ее увеличение до 75% [82].

Необходимо сделать существенную оговорку. Термин «услуги 3G» крайне расплывчат; из статистических данных зачастую невозможно понять, о какой же технологии идет речь. Например, сегодня самая массовая 3G-технология — cdma2000. Но она подразумевает плавный переход от 2G-стандарта IS-95a (cdma-One). При этом сохраняется полная обратная совместимость. Иными словами, оператор может постепенно дооснащать 3G-оборудованием свою CDMA-сеть, превращая ее в 3G-сеть cdma2000, а все «старые» телефоны как работали, так и будут продолжать работать. С технологией WCDMA ситуация принципиально иная, поэтому таких сетей гораздо меньше. Таким образом, очевидно, что, говоря о 3G-услугах, имеют в виду 3G-сети. Возможность же получения 3G-услуг зависит от наличия 3G-телефонов, которых в Европе да и во всем мире пока немного. Практически все развернутые 3G-сети имеют двухмодовый режим работы (2,5G/3G) и используются операторами для выявления числа абонентов, готовых оплачивать услуги, не связанные с передачей голоса и SMS. Подавляющее большинство абонентов таких сетей, хотя именуется «абонентами услуг 3G сетей», в реальности пользуются обычными 2,5G-телефонами. Отсюда берутся те многие миллионы абонентов 3G-сетей.

Тем не менее с внедрением 3G-технологий европейские и американские операторы испытывают проблемы. На лицензии потрачены огромные деньги, их надо возвращать. Как подсчитали аналитики из инвестиционного банка Schroder Salomon Smith Barney, лицензии окупятся только в том случае, если каждый европеец будет дополнительно приносить оператору связи не менее 500 евро ежегодно. В результате операторы все чаще приходят к выводу о необходимости объединения своих ресурсов в проектах по развертыванию 3G-сетей.

Гонконгская компания Hutchison Whampoa объявила о решении создатьковую связь 3G в Западной Европе. Она действует под учрежденной ей торговой маркой «3» (65% акций «3» принадлежит Hutchison, еще 20% владеет NTT DoCoMo) и к 2003 году уже вложила в проект порядка 16,7 млрд долл. Hutchison активно реализует свои планы создания 3G-сетей в Великобритании, Италии, Швеции, Австрии и других странах. Благодаря усилиям Hutchison к концу марта 2003 года первые 60 тыс. европейцев (50 тыс. в Италии и 10 тыс. в Великобритании) смогли оценить связь третьего поколения. Зона обслуживания абонентов сети UMTS в Великобритании составляет 70% территории (4,3 тыс. базовых станций), в Италии — 50% (3,2 тыс. базовых станций), соответственно 200 и 400 тыс. абонентов. Деятельность Hutchison в области 3G не ограничена Европой. Она запустила 3G-сеть и в Австралии, под тем же брендом «3», как и в Европе.

Китай пока не спешит внедрять сети 3G. Предполагалось, что в Китае рынок 3G-услуг сформируется к 2005 году. В частности, в докладе «Управление и развитие China Telecom в 2003 году» представитель компании ZTE (производитель телекоммуникационного оборудования) заявил, что оператор China Telecom выберет в качестве стандарта сотовой связи третьего поколения W-CDMA, а не TD-SCDMA. Этот оператор строит опытную тестовую сеть в семи китайских городах, включая Шанхай.

В России о технологиях 3G пока говорить рано. В то же время в нашей стране уже появилась вполне «3G-технология» — cdma2000 в диапазоне 450 МГц.

2.3.3 Технологии транковой радиосвязи

По определению, транкинг (trunking) — это метод равного доступа абонентов к общей для них группе каналов связи, при котором конкретный канал предоставляется абоненту для каждого сеанса связи заново. Иными словами, это метод доступа многих пользователей к общему каналному ресурсу. В этом смысле термин «транковые сети» можно трактовать очень широко — подавляющее большинство современных систем связи, включая обычную проводную телефонию, действуют именно по этому принципу. Однако, вопреки определению, под транковыми сетями связи понимают вполне определенный класс систем связи, а именно беспроводные корпоративные сети, использующие принцип свободного доступа абонентов к общему пулу каналов. Отметим, что русскоязычное определение сетей данного класса не устоялось — специалисты говорят и «транкинговый», и «транковый». Для определенности будем придерживаться последнего термина.

Расцвет транковых сетей пришелся на период начала 90-х годов, вплоть до становления систем сотовой связи. Возможно, первой в мире системой транковой связи была еще памятная многим отечественная система «Алтай». Может быть, она бы успешно работала и поныне, если бы массогабаритные и эксплуатационные характеристики ее радиотелефонов не перестали удовлетворять самых неприхотливых пользователей. Однако с развитием технологий сотовой связи на долю транковых систем остались весьма специфические области применения, где проявляются их основные достоинства. Это прежде всего малое время установления соединения (менее секунды), автономность сети (проще контролировать систему), возможность обеспечить высокую защищенность от несанкционированного доступа, как к каналному ресурсу, так и к передаваемым данным. Транковые сети эффективны там, где стоимость самой дорогой системы связи несоизмерима с возможными последствиями сбоев, случайных или преднамеренных, или несанкционированного доступа к данным. Речь идет о диспетчерских службах в энергетических и транспортных сетях, о силовых ведомствах, оперативных службах и т. п.

Виды транковых сетей связи

Исторически первые транковые системы пришли на смену корпоративным аналоговым радиосетям с жестким распределением каналов. Первые системы тран-

ковой связи были аналоговыми. По существу, они представляли собой системы аналоговой радиосвязи с каналами фиксированной ширины (в основном 12,5 или 25 кГц) и средствами предоставления доступа к этим каналам. Собственно, все стандарты (открытые и корпоративные) в области аналоговой транковой связи специфицируют именно систему сигнализации и самые общие требования к приемопередающей аппаратуре. Механизм непосредственной передачи в радиоканале находится «вне зоны их компетенции» — лишь бы не было межканальных помех. В наиболее простых аналоговых системах транковой связи абонентские станции представляли собой стандартные портативные радиостанции соответствующих диапазонов, оснащенные дополнительным транковым модулем.

По схеме построения транковые сети могут быть одно- и многозоновыми. В первом случае в систему входит одна базовая ретрансляционная станция (с одной зоной обслуживания). В многозональных сетях базовых станций несколько. При этом управление может быть как сосредоточено в центральном контроллере (централизованное управление), так и возложено на оборудование базовых станций (распределенное управление). Как правило, все транковые системы строятся по модульному принципу, т. е. в минимальной конфигурации это может быть одна базовая станция с двумя каналами. Впоследствии оборудование можно наращивать, вплоть до общенациональных сетей.

При свободном доступе к каналам возможны два основных механизма распределения ресурсов: когда поиском и распределением свободных каналов занимается базовая станция и когда абонентская станция сама находит свободный канал. Оба метода требуют передачи специальных команд — вызов, отбой, групповой вызов, аутентификация и т. п. Совокупность этих команд называют системой сигнализации. Управляющие сообщения могут передаваться как в специальном сигнальном канале, так и в рабочих каналах. В последнем случае сигнализация транслируется в собственно разговорном канале (в полосе 300–4000 Гц, внутриполосная сигнализация) либо вне его (в полосе 0–300 Гц, субтоновая сигнализация) — все как в обычной проводной телефонии.

Простейший и наиболее распространенный пример внутрисполосной сигнализации, широко распространенной в проводной телефонии и используемой в транковых системах, — это двухчастотная система DTMF, применяемая при «тоновом» наборе номера. Наиболее распространенная система внешеполосной сигнализации — CTCSS (Continuous tone controlled signaled system), описанная в британском стандарте MPT 1306. Она (в базовой конфигурации) представляет собой оговоренный набор из 32 частот в диапазоне 67–250,3 Гц.

Характерный пример распределенной системы транковой связи — это ставшая популярной в нашей стране к концу 90-х годов система SmarTrunk и ее дальнейшее развитие SmarTrunk II. Систему SmarTrunk впервые продемонстрировала компания Selectone (США) в 1992 году. Дальнейшим ее развитием занималась отпочковавшаяся в 1996 году фирма SmarTrunk Systems. SmarTrunk — это протокол, определяющий механизм доступа абонентов к 16 возможным аналоговым дуплексным каналам. Каждая станция, желавшая установить соединение, сканировала все доступные каналы и, обнаружив свободный, посылала центральной станции запрос на его захват. В результате процедура установления соединения занимала от 2 до 11 с. Система сигнализации — внутрисполосная, DTMF. Из-за

низкой надежности и защищенности последней и появилась усовершенствованная система SmarTrunk II с цифровой сигнализацией (BPSK). Все работа ведется через ретранслятор — свой для каждого дуплексного канала. Был реализован механизм аутентификации общего вызова, связи с городской телефонной сетью и т. п. Рабочий диапазон частот, вообще говоря, может быть произвольным, но наиболее распространенные диапазоны 146–174, 400–490, 300–344 и даже 800 МГц.

Наибольшее распространение получили аналоговые транковые системы на основе группы стандартов Министерства почты и телекоммуникаций Великобритании (MPT — Ministry of Posts and Telecommunications). Центральным в этой группе выступает стандарт MPT 1327 (A Signalling Standard for Trunked Private Land Mobile Radio Systems) — «Стандарт сигнализации частных наземных систем транковой подвижной радиосвязи». Он был опубликован в 1986 году и окончательно утвержден в 1987 году. Отметим, что хотя аналоговые транковые сети на основе стандартов MPT и называют (мы следуем этой традиции) «системы стандарта MPT 1327», однако требования к ним содержатся и в других стандартах группы MPT. В частности, спецификация MPT 1347 описывает интерфейс системного радиооборудования и контроллеры транковой связи, MPT 1343 — интерфейс абонентского радиооборудования, MPT 1317 — формат кода синхронизации и т. д. Поскольку аналоговые транковые сети стандарта MPT 1327 — самые распространенные в мире (кроме, пожалуй, США, где лидируют системы стандарта LTR), остановимся на нем несколько подробнее.

Аналоговые транковые сети стандарта MPT 1327

Разработчиком оборудования для первых систем MPT 1327 была фирма Fylde Microsystems. Позже был образован консорциум компаний-разработчиков, куда кроме Fylde вошли Tait Electronics, Philips, Motorola и Marconi.

Системы MPT 1327 изначально разрабатывались как многозоновые. В каждой зоне индивидуальные вызовы обрабатываются независимо. В случае обрыва междононных связей базовая станция продолжает работать, но уже без обработки междононных вызовов. Время соединения при внутризоновых вызовах не превышает 0,5 с, при междононных — 1–2,5 с.

Теоретически сеть на базе MPT 1327 может обслуживать до 1 036 800 абонентов и состоять из 1024 зон по 24 канала в каждой. Однако даже наиболее крупные сети MPT 1327 далеки от таких показателей. Так, одна из крупнейших сетей Chekker Network компании Deutsche Telekom в 1998 году насчитывала примерно 900 каналов в 160 зонах и обслуживала около 62 тыс. абонентов. В основном же транковые системы MPT 1327 включают 3–5 базовых станций с 4–8 каналами каждая и обслуживают 1–2 тыс. абонентов.

В системе MPT 1327 каждая станция имеет свой уникальный номер, присваиваемый производителем — ESN (Electronic Serial Number). Для работы в сети требуется начальная регистрация станции и последующая аутентификация на основании ранее зарегистрированного номера.

Характерная особенность сетей стандарта MPT 1327 — наличие выделенного канала управления, по которому происходит обмен управляющей информацией между базовыми и абонентскими станциями. Физически это один из частотных

каналов базовой станции. Остальные каналы предназначены для обмена речевой информацией и данными. Отметим, что при большой загрузке системы канал управления также можно использовать для передачи голоса или данных, хотя и с потерей некоторых функций. Канал управления может автоматически переноситься с одного частотного канала на другой (например, при появлении сильных помех).

Обмен сигнальной информацией происходит со скоростью 1200 бит/с посредством так называемой быстрой частотной манипуляции (FFSK — fast frequency-shift keying). Несущей с частотой 1200 Гц соответствует логическая «1», логическому «0» — 1800 Гц. Управляющая информация передается пакетами, содержащими интервал для включения передатчика (длительностью 5 мс), синхроследовательность (1010...10 — всего 16 бит), собственно информационное сообщение (64 бита) и последний бит согласования, значение которого зависит от последнего бита поля сообщения. Само сообщение состоит из кодового синхронизирующего синхрослова (C4D7₁₆ для канала управления, инверсное ему — для канала связи), признака адрес/данные (1/0), собственно информации (47 бит) и проверочной последовательности (16 бит). Для управления используется набор из 32 команд длиной 5 бит. Абонентские радиостанции постоянно принимают и анализируют сообщения в управляющем канале и, обнаружив в них собственный адрес, выполняют предписанные командой действия. Этот процесс происходит независимо от того, поддерживает абонентская станция в данный момент сеанс связи или нет. По управляющему каналу возможна передача не только командных сообщений, но и коротких (до 184 бит) и расширенных (736 бит) информационных сообщений (пейджинговые сообщения, данные GPS и т. п.).

Приемопередающее оборудование стандартом MPT 1327 не специфицируется. Для этого был разработан стандарт MPT 1347. В соответствии с ним интервал между несущими соседних каналов должен составлять 12,5 кГц, разнос дуплексных каналов (нисходящий/восходящий) — 8 МГц, рабочий диапазон восходящих/нисходящих каналов — 201, 2125–207, 4875 и 193, 2125–199, 4875 МГц соответственно. При передаче речи предусматривается фазовая модуляция, для передачи данных — FFSK.

Сети MPT 1327 позволяют передавать не только речевые сообщения, но и данные со скоростями до 1200 бит/с. Особенность работы сетей в режиме прозрачности для цифровых данных описывает протокол MAP 27 (Mobile Access Protocol for MPT 1327 equipment).

Отметим, что на основе стандарта MPT 1327 делались попытки построить и цифровые системы (например, ACCESSNET-D компании Rohde & Schwarz), однако особого успеха они не имели.

Система TETRA — цифровая система транковой связи

Естественным развитием транковой связи стало появление цифровых систем, т. е. систем, где и служебная информация, и голосовые каналы передаются в цифровой форме. Основное достоинство цифровых систем применительно к транковой связи — это существенно более эффективное использование частотного ресурса. В самом деле, благодаря цифровым алгоритмам сжатия речи, эффек-

тивная ширина голосового канала снижается до 6,25 кГц и ниже, что по крайней мере в два-четыре раза меньше, чем в аналоговых системах. Самой первой цифровой системой транковой связи была система EDACS (Enhanced Digital Access Communications System) шведской компании Ericsson. Позже появились системы TETRA, APCO 25, TETRAPOL (компания Matra), iDEN (Motorola), ACCESNET-D (Rohde & Schwarz) и некоторые другие. Судьбы их были различны, отметим лишь, что только первые две из этого списка базировались на открытых стандартах и только в TETRA использовался механизм временного разделения каналов (TDM). Во всех остальных разделение каналов частотное (FDM). Возможно, именно это и предопределило определенный успех стандарта TETRA в Европе, включая и Россию.

В декабре 1994 года по инициативе Motorola и Nokia представители 14 крупнейших производителей, среди которых были Philips, Alcatel, Ericsson, GEC-Marconi, подписали Меморандум о взаимопонимании в создании единого европейского стандарта транковой связи TETRA. Через год работу над новым стандартом продолжила рабочая группа Европейского института телекоммуникационных стандартов (ETSI). А еще через два года стало очевидно: несмотря на наличие конкурирующих систем, стандарт TETRA все более укрепляет свои позиции, приобретая общемировое значение.

До апреля 1997 года аббревиатура TETRA означала Трансевропейское транковое радио (Trans-European Trunked RAdio). Однако в связи с огромным интересом, который проявляют к стандарту в Южной Америке, Африке и Азии, территория его действия уже не ограничивается только Европой. Поэтому сегодня TETRA расшифровывают как Наземное транковое радио (TErrestrial Trunked RAdio).

С самого начала стандарт TETRA разрабатывался как транковая система, эффективно и экономично поддерживающая совместное использование сети несколькими организациями, обеспечивая секретность и защищенность информации. Особое внимание в стандарте уделено интересам служб безопасности. Существенно, что TETRA — открытый стандарт. Это означает, что соответствующее стандарту оборудование различных производителей абсолютно совместимо. Доступ к спецификациям TETRA свободен для всех заинтересованных сторон, вступивших в ассоциацию «Меморандум о взаимопонимании и содействии стандарту TETRA» (MoU TETRA). Ассоциация объединяет производителей, разработчиков программного обеспечения, испытательные лаборатории и пользователей из 19 стран мира.

Стандарт TETRA не описывает всю сеть детально, как, скажем, GSM. В него входят спецификации беспроводного интерфейса, интерфейсов между сетью TETRA и цифровой сетью с интеграцией услуг (ISDN), телефонной сетью общего пользования, сетью передачи данных общего пользования, учрежденческими АТС и т. п. В стандарт включено описание всех основных и дополнительных услуг, предоставляемых сетью. Специфицированы также интерфейсы локального и внешнего центрального управления сетью.

Технические особенности системы TETRA

TETRA представляет собой транковую систему связи, основанную на технических решениях и рекомендациях стандарта GSM. Радиоканалы разнесены с ин-

тервалом 25 кГц. Дуплексный разнос радиоканалов для передачи и приема — 10 МГц. В системе использована технология временного разделения каналов (TDMA) — в одной частотной полосе последовательно отводится время для передачи четырех логических каналов (рис. 2.46). Эти четыре канальных интервала (слота) образуют TDMA-кадры. Сообщения передаются мультикадрами продолжительностью 1,02 с. Мультикадр содержит 18 TDMA-кадров, один из которых — контрольный. TDMA-кадр состоит из четырех временных пакетов по 510 бит каждый. 432 бита в пакете (два блока по 216 бит) относятся к информационному сообщению. В середине пакета — синхропоследовательность SYNCH для синхронизации и обучения адаптивного канального эквалайзера в приемнике. Кроме того, пакеты в восходящем канале содержат интервал PA (Power Amplifier), предназначенный для установки мощности излучения. В этом случае в конце пакета размещается защитный интервал GP (Guard period) длительно-стью 0,167 мс (эквивалентно времени передачи 6 бит), исключаящий перекрытие соседних пакетов.

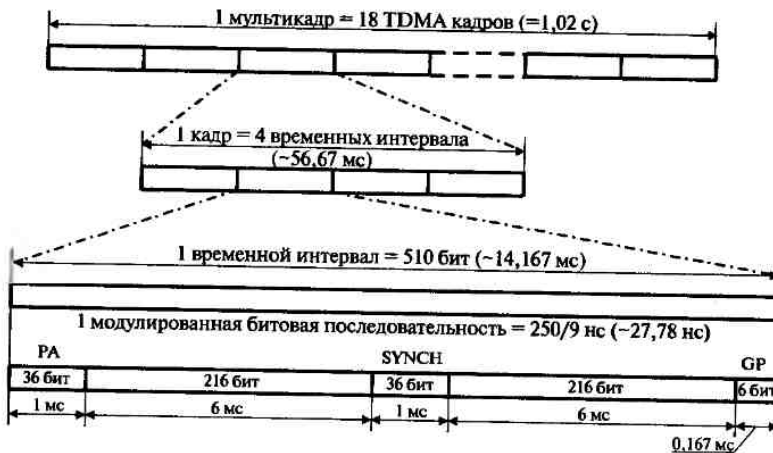


Рис. 2.46. Протокол информационного обмена системы TETRA

Передачу четырех речевых каналов в полосе 25 кГц обеспечивает низкоскоростной кодер речи с алгоритмом CELP. Скорость потока на выходе кодера 4,8 кбит/с. Общая же скорость цифрового потока данных после кодирования и включения служебной информации — 36 кбит/с. Однако благодаря применяемой в системе $\pi/4$ -DQPSK-модуляции (дифференциальная квадратурная фазовая модуляция) частота модуляции каждой несущей вдвое меньше — 18 кбит/с.

TETRA допускает прямую связь между радиостанциями без ретранслятора и расширение зоны радиопокрытия с помощью мобильных станций. Системы TETRA способны работать в широком диапазоне частот — от 60 до более чем 1000 МГц. За службами безопасности уже закреплены диапазоны 380–385/390–395 МГц. Для коммерческих организаций выделяются полосы 410–430/450–470 МГц и 870–876/915–921 МГц.

Разработаны два стандарта системы: TETRA PDO для передачи данных и TETRA V+D, поддерживающий как передачу данных, так и речевой обмен. Далее будем рассматривать последний стандарт.

Архитектура системы определяется условиями применения. Благодаря модульному принципу построения она может быть реализована с разными иерархическими уровнями и географической протяженностью (от локальной до национальной сети). Функции управления базой данных и коммутацией распределяются по всей сети. Это гарантирует быструю передачу вызова и локальную работоспособность даже при потере связи с отдельными ее элементами.

На национальном или региональном уровне структура сети строится на основе сравнительно небольших подсетей TETRA, соединенных друг с другом (рис. 2.47). При этом возможно централизованное управление всеми подсетями. Кроме того, для соединения сетей TETRA применяется и стандартный межсистемный интерфейс (ISI). Структура подсети зависит от трафика и требований к эффективности и доступности. Поэтому в одних случаях она может быть простой, в других, когда элементы сети связаны друг с другом многочисленными каналами связи, сравнительно сложной (рис. 2.48). Если не требуется резервирования каналов, оптимальна конфигурация звезды (рис. 2.49). При использовании линейных трактов наилучшее решение — структура цепи (рис. 2.50). В этом случае каждый модуль устройства управления базовой станцией обеспечивает как требуемую дальность радиосвязи, так и локальный доступ к внешним сетям (телефонные сети общего пользования, учрежденческие АТС). Простейшая конфигурация сети включает лишь один такой модуль.

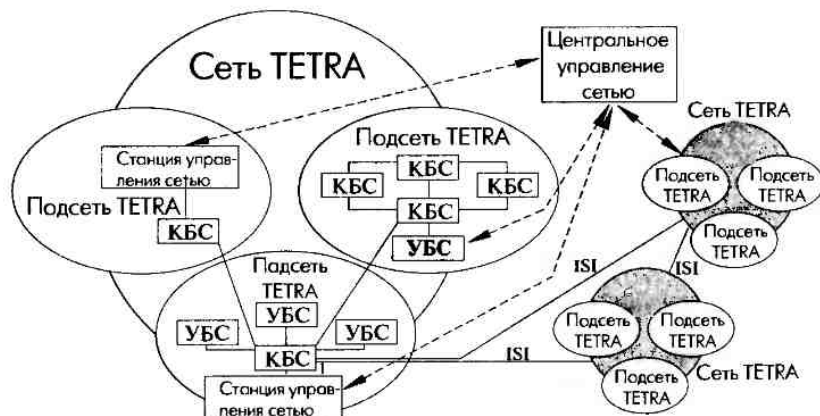


Рис. 2.47. Структура глобальной сети TETRA: КБС — контроллер базовой станции, УБС — блок управления базовой станцией, ISI — межсистемный интерфейс

Как уже отмечалось, система поддерживает трансляцию речи и данных. При этом речь и данные могут передаваться с одного терминала одновременно.

Речь передается в трех режимах:

- 1) коммутируемое двухточечное соединение между стационарными терминалами и мобильными абонентами; вызовы могут быть переданы в телефонную сеть общего пользования или во внешние учрежденческие АТС; связь — дуплексная;

- 2) коммутируемые многопунктовые двунаправленные соединения между вызывающей стороной и несколькими вызываемыми абонентами; связь — симплексная;
- 3) широковещательная передача речи (симплексная) от вызывающей стороны к нескольким вызываемым абонентам.

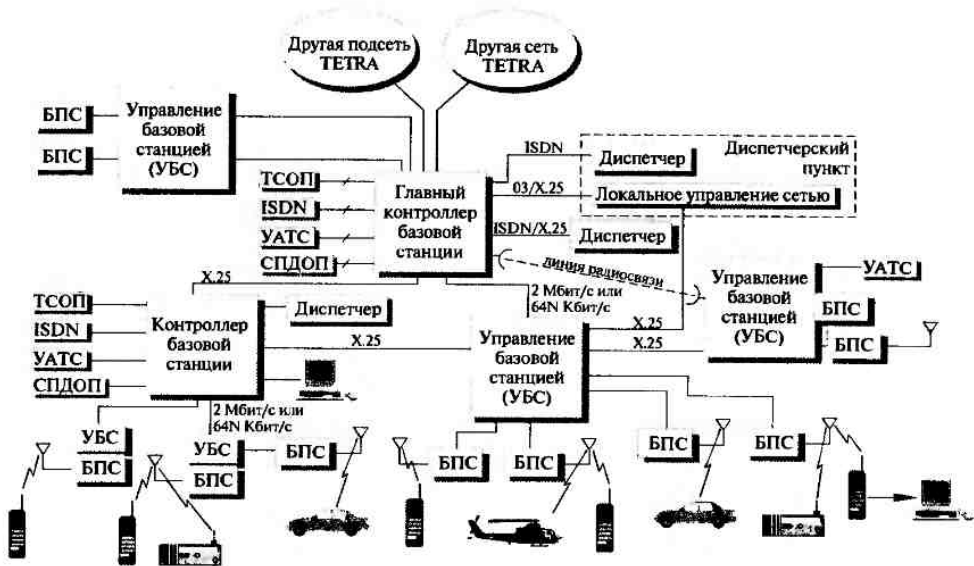


Рис. 2.48. Возможная конфигурация подсети TETRA: ТСОП — телефонная сеть общего пользования, СПДОП — сеть передачи данных общего пользования, БПС — базовая приемопередающая станция

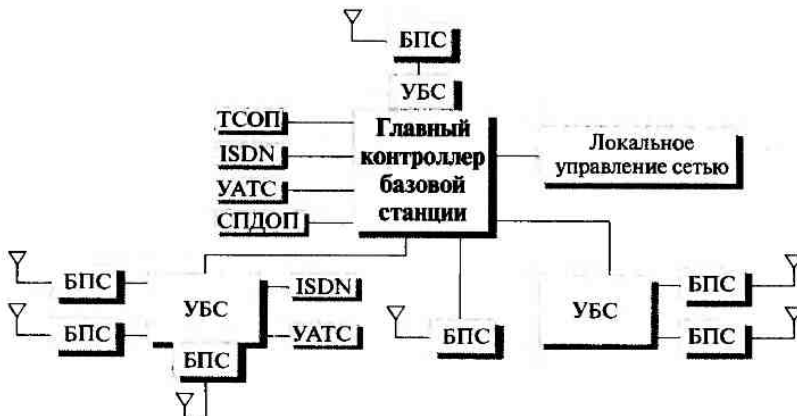


Рис. 2.49. Подсеть TETRA с конфигурацией типа «звезда»

Режимы передачи данных обеспечивают доступ к централизованным базам данных, передачу факсимильных сообщений, неподвижных изображений, видео-сигналов и т. д.

Передача данных с коммутацией цепей аналогична трансляции речи — те же режимы (двухточечная, многопунктовая, широковещательная передача) и характеристики (дуплексная, симплексная связь; обычный и защищенный трафик с несколькими скоростями передачи). Скорость обмена определяется используемым классом защиты от ошибок и числом временных интервалов, выделенных для связи (табл. 2.8). Число выделенных временных интервалов задается при вызове и может меняться в процессе передачи.



Рис. 2.50. Подсеть TETRA с конфигурацией типа «цепь»

Таблица 2.8. Связь скорости передачи в сети с уровнем защиты и числом интервалов

Уровень защиты	Скорости передачи, кбит/с			
	1 интервал	2 интервала	3 интервала	4 интервала
Защита отсутствует	7,2	14,4	21,6	28,8
Низкий уровень	4,8	9,6	14,4	19,2
Высокий уровень	2,4	4,8	7,2	9,6

Коммутируемые пакеты данных транслируются по виртуальным цепям или в виде датаграмм. В первом случае возможны только двухточечные соединения, а во втором — многоточечные соединения и широковещательная передача. Пользователи системы могут подключаться и к внешним сетям пакетной передачи данных. Короткие сообщения (до 2048 бит) оперативно передаются независимо от трансляции речи и данных.

Каналы передачи данных выделяются в соответствии с тремя сценариями.

1. Канал присваивается в начале сеанса связи и освобождается по его завершении. Обычно данный сценарий используется для дуплексной связи или при резервировании канала в случае его непрерывной эксплуатации определенной группой пользователей (например, режим открытого канала). Его недостаток заключается в том, что канал нельзя задействовать при динамическом распределении ресурсов (система менее эффективна с точки зрения управления радиоресурсами).

2. Канал присваивается только на время одной транзакции (интервал передачи/прием), после чего он освобождается. Для следующей транзакции с тем же вызовом может быть выделен новый канал. Этот метод эффективен при групповых вызовах, когда для различных вызовов совместно используется вся совокупность каналов.
3. В режиме квазипередачи данных, в отличие от предыдущих сценариев, канал освобождается с некоторой задержкой. И если следующая транзакция началась во время задержки, она задействует тот же канал, что и предыдущая. Достоинство метода — в уменьшении необходимых управляющих сигналов в сети.

В режиме с открытым каналом группа пользователей организует соединение по схеме «точка-многоточка» без какой-либо установочной процедуры, в том числе без задержки. Присоединившись к группе, можно воспользоваться этим каналом в любой момент. Связь — двухчастотная, симплексная.

При непосредственной передаче между терминалами устанавливаются двух- и многоточечные соединения по радиоканалам, не связанным с каналом управления сетью, без передачи сигналов через базовые приемопередающие станции. Радиостанции работают в симплексном режиме на одной частоте. В этом режиме отдельные мобильные станции могут выполнять функции ретрансляции (повторителя) для увеличения зоны действия других станций, а также служить шлюзом между мобильной станцией, работающей в режиме непосредственной передачи данных, и сетью TETRA.

Пользователи сети регистрируются в соответствии с принадлежностью к определенной зоне, обслуживаемой несколькими базовыми приемопередающими станциями. В пределах данной зоны абоненты могут свободно перемещаться и устанавливать связь друг с другом. Протяженность зоны обслуживания конкретного пользователя определяется его потребностями и привилегиями и может колебаться от радиуса действия базовой станции до всей сети. Различные зоны предоставляют одинаковые услуги, которые могут поддерживать все радиотерминалы. Каждый пользователь имеет собственный «профиль» — список доступных услуг, хранящийся в его терминале и сети.

Если в процессе соединения мобильная станция регистрирует ухудшение условий связи, она проверяет возможность установления связи в соседних ячейках (повторный выбор ячейки) и посылает в сеть запрос на новый радиоканал (если вызов не потерян). Для повторного выбора ячеек предусмотрено три типа процедур. Процедуру первого типа можно сравнить с «бесшовным» переключением в системе сотовой связи, подобной GSM. В этом случае мобильная станция должна обладать информацией о новой ячейке и о распределении ее каналов (при необходимости запрос для регистрации передается заранее посредством старой ячейки). При процедуре второго типа мобильная станция владеет информацией о новой ячейке, но не о распределении ее каналов. И наконец, при процедуре третьего типа мобильная станция ничего не знает о новой ячейке. Однако она посылает старой ячейке сообщение о предстоящей замене на новую.

Дополнительные услуги

Система TETRA предоставляет пользователям немало дополнительных услуг. Часть из них введена в стандарт по заявке Ассоциации европейской полиции (Schengen Group), которая тесно сотрудничает с техническим комитетом ETSI, отвечающим за разработку TETRA. Так, специально для нужд служб общественной безопасности поддерживается режим, по которому вызовы поступают абонентам только при наличии санкции диспетчера. Существует система приоритетов доступа, действующая при перегруженности сети. Приоритеты присваиваются и вызовам (в соответствии с приоритетом абонента). Группы пользователей могут создаваться и модифицироваться динамически. При групповом вызове предусмотрена идентификация вызывающей стороны. Кроме того, допускается избирательное прослушивание терминалов.

Поддерживаются и другие услуги. Пользователю предоставляется возможность задавать зоны вызовов — вызывной сигнал дойдет только до абонентов в пределах выбранных зон. По его желанию на терминале вызываемого абонента отображается идентификатор вызывающей стороны и наоборот. В ходе сеанса связи пользователь получает информацию о попытке соединения с ним, а вызываемому абоненту сообщается о том, когда соединение возможно. Постоянно или при необходимости (номер абонента занят, не отвечает или находится за пределами зоны обслуживания) вызовы перенаправляются по другому номеру или по первому доступному номеру из списка абонентов, предоставленного пользователем. Предусмотрена возможность принимать вызовы и в процессе разговора. С помощью специальной опции разрешаются либо блокируются определенные категории входящих и исходящих вызовов, например междугородних. Немаловажно и то, что абонентам поступает информация об оплате услуг.

Защита информации

Защита информации — важнейший аспект построения системы TETRA, поскольку одной из основных групп пользователей являются службы общественной безопасности, для которых высокий уровень защиты — обязательное требование. Система защиты TETRA основана на трех принципах: структурированность, открытость и применение хорошо проверенных методов. Структурный подход к разработке и построению системы позволил сформулировать требования к защите (ETSI TETRA Technical Requirements Specification ETR 086-3) еще на этапе проектирования. На базе этих требований разработаны специальные функции и механизмы защиты, вошедшие в протоколы TETRA и составляющие с системой единое целое. Описание механизмов и алгоритмов системы защиты TETRA доступно всем желающим. Иными словами, защищенность информации не зависит от сохранения в секрете способа шифрования.

В стандарт вошли только хорошо проверенные методы защиты, в первую очередь из систем GSM и DECT. Это механизмы аутентификации мобильного терминала, обеспечение конфиденциальности радиоканала (GSM), а также взаимная аутентификация терминала с сетью и функции управления ключами кодирования (DECT). Однако в стандарт включен и ряд других функций, обеспечиваю-

щих совместное соблюдение требований как GSM, так и DECT и не препятствующих быстрому установлению соединений и мобильности пользователей.

При рассмотрении средств защиты TETRA следует четко различать три класса функций:

- 1) механизмы защиты выполняют основную работу по обеспечению безопасности системы, реализуя такие функции, как идентификация пользователей и конфиденциальность информации. Это основные структурные блоки в системе защиты;
- 2) функции управления защитой управляют отдельными механизмами защиты, контролируя их непротиворечивость и корректность при взаимодействии разных сетей;
- 3) стандартные криптографические алгоритмы — суть стандартизированные математические функции, обычно используемые в комбинации с ключами. Они реализуют адекватный уровень защиты в соответствии с применяемыми защитными механизмами. Эти алгоритмы, входящие в систему TETRA как опции, обеспечивают взаимодействие различных сетей.

Средства обеспечения законности снимают противоречия между национальными законодательными требованиями, которым должна соответствовать система связи, и системой защиты. Национальные законодательные акты соблюдаются без изменения и ухудшения работы внутренней защиты TETRA.

Один из основных элементов системы управления защитой — ключевые последовательности (ключи). Они применяются при процедурах аутентификации и шифрования информации.

Аутентификационный ключ используется для взаимного опознавания мобильного терминала и базовой станции. Применяются три вида таких ключей:

- 1) пользовательский аутентификационный ключ (UAK) длиной 128 бит, хранящийся в памяти мобильной станции или Smart-карты;
- 2) аутентификационный код, вводимый пользователем вручную;
- 3) комбинация UAK и вводимого пользователем персонального идентификационного номера (PIN-кода).

При передаче сообщений защитой от несанкционированного прослушивания служит шифрование. Шифровальные ключи можно формировать, распределять, выбирать и отменять при установлении связи между абонентами. Используют четыре вида шифровальных ключей:

- 1) выведенные ключи применяются для двухточечной связи и генерируются при процедуре аутентификации;
- 2) статические ключи представляют собой набор фиксированных кодовых последовательностей (до 32), которые могут использоваться без предварительной аутентификации;

- 3) общий ключ используется для шифрования при передаче групповых вызовов, формируется с использованием выведенного ключа для каждой мобильной станции, действует в строго определенной зоне и периодически изменяется;
- 4) групповой ключ, связанный с определенной группой пользователей, генерируется системой и передается мобильным станциям заданной группы. Он применяется для шифрования вызовов группы как в исходном виде, так и после модификации посредством общего ключа.

2.3.4 Широковещательные сети — цифровое телевидение

Краткий исторический экскурс

Телевидение — это одна из важнейших областей применения технологий беспроводной высокоскоростной передачи данных. Первой ТВ-трансляцией по СПИ считают передачу телевизионного изображения президента США Герберта Гувера по телефонным линиям из Вашингтона в Манхеттен в 1927 году, организованную компанией Bell System.

Телевидение — это еще и одна из наиболее «древних» областей электронной техники, по возрасту не уступающая радио. Работы в этом направлении начались еще до появления беспроводной связи. Основой для многих систем раннего, так называемого оптико-механического телевидения был созданный в 1884 году и запатентованный в 1895-м немецким ученым Паулем Нипковым вращающийся перфорированный «развертывающий диск» из селеновых ячеек, раскладывающий изображение на элементы. В 20-е и в начале 30-х годов во всех странах работы в области телевидения, вплоть до попыток создания сетей вещания и цветного телевидения, были связаны именно с этим устройством.

Примечательно, что история создания и развития телевидения во всем мире неразрывно связана с отечественными специалистами. Так, сам термин «телевидение» ввел в оборот петербургский инженер Константин Дмитриевич Перский в 1900 году на научной конференции в Париже. Определяющими для электронного телевидения стали работы профессора Петербургского технологического института Бориса Львовича Розинга с зеркальными способами оптико-механической развертки (1907–1911 годы). Он использовал катодно-лучевую трубку англичанина Уильяма Крукса, причем механизм развертки принципиально не отличался от современного. Однако несовершенство первых электронных ламп и отсутствие усилителей не позволяли в то время получить приемлемое изображение. Но самое главное — Б.Л. Розинг основал научную школу, из которой вышло немало будущих создателей телевидения. Судьба Б.Л. Розинга типична для многих выдающихся россиян, оставшихся в России в 20-е годы, — арест в 1931 году и смерть в ссылке в Архангельске.

Время с 1920 до 1930 год — период бума в области систем механического телевидения малой строчности (порядка 30 строк). В США, в Европе, в СССР создаются так называемые малострочные ТВ-системы. Появляются первые приемники и видеокамеры. В 1931 году в США насчитывалось уже 21 экспериментальная ТВ-станция, порядка 8 тыс. малострочных приемников. В Великобри-

тании к 1931 году было продано около тысячи ТВ-приемников Джона Лоджи Берда с разверткой в 30 строк. Постепенно число строк механических ТВ-систем перевалило сотню, но в середине 30-х все подобные системы начало вытеснять электронное телевидение.

Примечательно, что одну из наиболее совершенных систем с механической разверткой разработал гениальный российский инженер и профессиональный музыкант Лев Сергеевич Термен. Массовой публике он известен как создатель первого электромузыкального инструмента «терменовокс». Однако еще в 1926 году он построил установку телевизионного наблюдения с разрешением 64 строки, экраном $1,5 \times 1,5$ м и системой передачи изображения на 50 м («Радиолобитель», 1927, № 1). Л.С. Термен продолжил работы над совершенствованием ТВ-системы наблюдения для пограничных войск с очень жесткими тактико-техническим требованиями. Так, число строк развертки превышало сотню, аппаратура должна была работать в полевых условиях и т. п. С помощью этого устройства К.Е. Ворошилов в 1927 году наблюдал из своего кабинета за посетителями во дворе здания Наркомата обороны. К сожалению, продолжить деятельность в ТВ Л.С. Термен не смог — он стал активно сотрудничать с советской разведкой, в конце 20-х начинается его музыкальное турне с терменовоксом в Европе, закончившееся в США. Там он, по-видимому, продолжая разведывательную деятельность, организовал фирму Teletouch по производству терменовокса и различных его модификаций, создал первое цветомузыкальное устройство, самое главное — охранную систему (использовавшуюся даже в Форт-Ноксе). Стал миллионером. В 1938 году был отозван (или был вынужден уехать) в СССР. Там его ждал лагерь, знакомство с С.П. Королевым на колымских золотоносных приисках, знаменитая «туполевская шарага» в Омске, работа над системой радиоуправления беспилотных самолетов и подслушивающими устройствами. Умер Лев Сергеевич в 1993 году в возрасте 95 лет в нищете и безвестности (в США о нем помнят до сих пор, снят документальный фильм «Электронная одиссея Льва Термена»).

Иначе сложилась судьба двух других россиян, которых в США называют «отцами телевидения». Речь идет о В. Зворыкине и Д. Сарнове. Давид Абрамович Сарнов, иммигрант из-под Минска, стал одним из ведущих инженеров у Г. Маркони. В 1912 году прославился тем, что принял сигнал SOS с «Титаника» и в течение трех суток поддерживал радиосвязь с судами, спасшими часть пассажиров. В 1930 году стал президентом основанной Маркони компании Radio Corporation of America (RCA) и ее дочерней радиосети National Broadcasting Company (NBC). Его роль в развитии телевидения в США столь велика, что в 1944 году на первом собрании только что созданного Общества телевизионных вещателей США именно Сарнову, тогда бригадному генералу, был преподнесен титул «отца телевидения». Примечательно, что Д. Сарнов был личным советником практически всех президентов США, начиная с Вудро Вильсона.

В 1929 году Д. Сарнов пригласил на работу в RCA Владимира Кузьмича Зворыкина — известного к тому времени в США специалиста в области создания электронных трубок, сотрудника компании «Вестингауз». За свою жизнь (1889–1982) В.К. Зворыкин стал обладателем 120 патентов (последние — в области управляемых ракет, компьютеров для предсказания погоды, электронных микроскопов и т. д.). В 1912 году Зворыкин закончил Петербургский технологи-

ческий институт, где познакомился с Б.Л. Розингом и принимал участие в его работах над системами электронного телевидения. В 1919 году эмигрировал в США, уже год сотрудничая с фирмой «Вестингауз». В США Зворыкин возобновил работы над телетехникой. Им разработан кинескоп. В 1931 году в RCA он создает иконоскоп, преобразующий оптическое изображение объекта в электрический сигнал, и начинает работы над созданием телевизионной системы в целом. Примечательно, что одновременно с ним в СССР аналогичное изобретение сделано группой сотрудников Всесоюзного электротехнического института под руководством Семена Исидоровича Катаева (приоритет заявки Катаева — на два месяца раньше, чем в патенте Зворыкина).

Регулярное электронное ТВ-вещание впервые началось в Европе. В Англии это произошло 2 ноября 1936 года компанией BBC. Использовались одновременно две системы: механического ТВ Берда (240 строк) и электронные ТВ-системы компании «ЭМИ-Маркони» (405 строк). Последняя использовала практически скопированный иконоскоп Зворыкина, назвав его «эмитрон». Это связано с тем, что руководитель телевизионной группы компании «ЭМИ-Маркони» Айзек Шенберг учился в Петербургском технологическом институте у Розинга, поддерживал тесный контакт с Зворыкиным. Между RCA и «ЭМИ-Маркони» было заключено соглашение об обмене научной информацией.

В Германии в 1931 году Манфред фон Арденне продемонстрировал работу электронной системы телевидения со стандартом 100 строк. Тогда же появились и первые телевизоры с электронным кинескопом. С 15 января 1936 года берлинский телецентр начал ежедневные трансляции с разрешением 180 строк. Олимпиада в Берлине в августе 1936 года транслировалась по 7–8 часов в сутки. Помимо индивидуальных телевизоров, передачи можно было смотреть в 25 просмотровых залах Берлина, а также в Гамбурге, куда был проложен кабель. За Берлинской Олимпиадой по телевизору следило в общей сложности 150 тыс. человек.

В Соединенных Штатах вещание электронного ТВ началось 30 апреля 1939 года компанией NBC с выступлений президента США Ф.Д. Рузвельта. В день начала передач в Нью-Йорке было всего 200 телевизоров. К концу 1941 года в США насчитывалось уже порядка 7,5 тыс. цифровых телевизоров.

В СССР регулярное ТВ-вещание началось 10 марта 1939 года на базе американского оборудования с разрешением 343 строки. С 16 июня 1949 года началось вещание в современном стандарте 625 строк.

Сегодня мы переживаем очередную технологическую революцию в телевидении — начало массового распространения цифрового телевизионного вещания. Цифровое телевидение (ЦТВ) — это принципиально новые возможности, интерактивность, среда доставки мультимедийного трафика. Поэтому изменение формата ТВ-вещания — не просто сложная техническая задача. Это — серьезнейший фактор, действующий в экономическом и социальном плане в общемировом масштабе.

Системы цифрового телевидения

Что дает цифровое телевидение? До массового зрителя практически без искажений доходит сигнал студийного качества, исчезают помехи, характерные для

аналогового вещания. Появляется возможность передавать видеоизображение телевидения высокой четкости (ТВЧ, HDTV) с числом строк развертки 720, 1080 (соотношение строк и столбцов 9:16) и выше против стандартных сегодня 480–625 строк (формат 3:4). Однако увидеть эту высокую четкость можно лишь на экране ТВ-приемника с соответствующими характеристиками кинескопа. Поэтому не менее важно, что ЦТВ позволяет гораздо эффективнее использовать спектральный диапазон — в полосе одного аналогового ТВ-канала можно формировать несколько цифровых, на порядок возрастает число одновременно доступных ТВ-программ. Наконец, развитие средств доставки цифрового телевизионного сигнала — важный шаг к столь интенсивно прорабатываемому сегодня «телевидению по запросу», когда зритель выбирает уже не канал, а непосредственно тот фильм или передачу, которые хочет смотреть.

Выделяют четыре основных механизма передачи ТВ-трафика конечным потребителям — кабельный, спутниковый, наземный (эфирное вещание) и так называемый сотовый. Последний метод реализуют высокочастотные системы с труднопроизносимыми аббревиатурами: MMDS (Multichannel Microwave Distribution System), LMDS (Local Microwave Distribution System), MWS (Multimedia Wireless System), работающие (в России) в диапазонах 2,5–2,7; 27,5–29,5 и 40,5–43,5 ГГц соответственно. Нас системы ЦТВ будут интересовать как пример широкополосных БСПИ, характерная особенность которых — выраженная асимметрия трафика. Более того, с появлением стандартов группы IEEE 802.16 становится возможным изготавливать и совместимые системы сотового телевидения. Российский зритель знаком (или вскоре познакомится) со спутниковым и кабельным ЦТВ. Однако стратегический вопрос развития телевидения в государстве — сеть наземного вещания. Остановимся на ней подробнее.

В области массового вещания столкнулись два стандарта передачи сигнала: ATSC (Advanced Television Systems Committee, США) и DVB (Digital Video Broadcasting, Европа). Отдельно стоит Япония со стандартом ISDB (Integrated Services Digital Broadcasting). Противостояние американского и европейского стандартов, по крайней мере для России, — уже вопрос прошлого, поскольку выбор в пользу DVB сделан, по всей видимости, окончательно. Эти стандарты аналогичны в том смысле, что оба ориентированы на передачу видео- и аудиоданных, кодированных и компрессированных посредством MPEG-2 [66]. Звук может кодироваться с помощью других алгоритмов, например Dolby AC-3. Поэтому качество картинки мало зависит от выбора одного или другого метода модуляции, если она успешно принята приемником.

В США при выборе стандарта цифрового телевизионного вещания основное внимание уделяли повышению качества изображения. При этом увеличение числа каналов со стандартным качеством изображения не рассматривалось. Очевидно, что изображение с высоким разрешением проявит себя на телевизорах с большим экраном — не меньше 32 дюймов, еще лучше — 55–60 дюймов. ТВ-приемники данного класса стоят соответственно 2–3 и 5–10 тыс. долл., т.е. гораздо дороже привычных телевизоров.

В Европе при разработке цифровой телевизионной системы DVB учли печальный опыт десятилетней давности, связанный с развертыванием коммерческого телевизионного вещания высокой четкости, в основном аналоговой ТВЧ-системы

MAC (Multiplexed analog components). Поэтому основное внимание уделили увеличению числа каналов со стандартной разрешающей способностью. Это позволило начать переход к ЦТВ с выпуска относительно дешевых (400–600 долл.) приставок для уже имеющихся телевизионных приемников. Кроме того, европейский стандарт рассматривается как основа для единой телекоммуникационной системы, ориентированной на передачу данных самой различной природы, отсюда требования к низкой вероятности ошибок. В то же время в ATSC отсутствуют опция передачи дополнительных данных.

ATSC ориентирован на трансляцию в полосе шириной 6 МГц (стандартная ширина канала американской системы аналогового ТВ NTSC) одного потока со скоростью 19,28 Мбит/с при наземном вещании и двух таких потоков — в сетях кабельного телевидения. DVB гораздо гибче: в стандартной полосе 8 МГц он обеспечивает выбор скорости в диапазоне от 4,98 до 31,67 Мбит/с (возможна работа в регионах с каналами шириной 6 и 7 МГц). Соответственно изменяется и число ТВ-программ в этой полосе — от 16 до 2, причем возможна одновременная трансляция программы с низким разрешением, но высокой надежностью и с высоким разрешением при пониженной надежности приема. Уникальная особенность DVB — это мобильность приемника, он может перемещаться со скоростью до 300 км/ч — это поезда, междугородный пассажирский автотранспорт, мобильные службы (скорая помощь, полиция) и т. п.

Однако самое существенное — это надежность доставки сигнала, в чем ATSC уступает DVB. Причиной этого является принятая в ATSC многопозиционная амплитудная модуляция с частично подавленной боковой полосой 8-VSB (vestigial-sideband modulation system for broadcast). В целом 8-VSB можно назвать вершиной технической реализации систем амплитудной модуляции с подавлением боковой полосы. Но по сравнению с DVB в реальных условиях этот механизм проигрывает по надежности, гибкости, требованиям к приемным антеннам. Рассмотрим конкурирующие стандарты подробнее.

Стандарт ATSC

ATSC разрабатывался с учетом того, что какое-то время в одном диапазоне будут транслироваться сигналы как в стандарте ATSC, так и в действующем в США аналоговом стандарте NTSC. Постепенно NTSC должен быть вытеснен, и весь занятый аналоговым вещанием диапазон станет цифровым. В отличие от ATSC спектр NTSC достаточно неравномерен. Поэтому для борьбы с интерференционными помехами в ATSC предусмотрен режекторный гребенчатый фильтр с подавлением основных составляющих сигнала NTSC (яркость, цвет и звук). Однако данный фильтр вдвое ухудшает для приемника соотношение сигнал/шум, поэтому он включается, только если приемник обнаруживает в эфире NTSC-сигнал.

Очень кратко рассмотрим систему VSB, принятую в ATSC. VSB — это амплитудная модуляция с подавлением нижней боковой полосы и частичным подавлением несущей. Уровней модуляции может быть от 2 (2-VSB) до 16 (16-VSB), они располагаются симметрично относительно 0. Так, амплитуда сигнала в 8-VSB принимает значения, пропорциональные целым числам от -7 до 7 с шагом 2.

Для кабельного вещания принята наиболее быстрая система 16-VSB, для наземного — 8T-VSB с предварительным решетчатым (сверточным) кодированием со скоростью $2/3$ (T — от trellis, решетка).

На вход VSB-системы поступают пакеты MPEG-2 объемом 188 байт (рис. 2.51). Из них удаляется байт синхронизации, к оставшимся 187 байтам добавляется 20 проверочных байтов кода Рида–Соломона. Дополнительно включенный в 8T-VSB решетчатый кодер превращает каждые 2 бита получившегося 207-байтового слова в 3 бита, т. е. в один символ. Очевидно, что скорость передачи данных пропорциональна двоичному логарифму от числа уровней амплитудной модуляции (число бит на символ), однако чем больше уровней, тем ниже помехозащищенность. Частота следования символов в VSB — 10,76 МГц. В результате скорость в системе 16-VSB, где каждым 4 битам соответствует один символ, в два раза выше (38,6 Мбит/с), чем в 8T-VSB (19,3 Мбит/с), поскольку для передачи одного пакета требуется вдвое меньше символов. Однако и пороговый уровень соотношения сигнал/шум у данных систем соответственно 28,3 и 14,9 дБ.

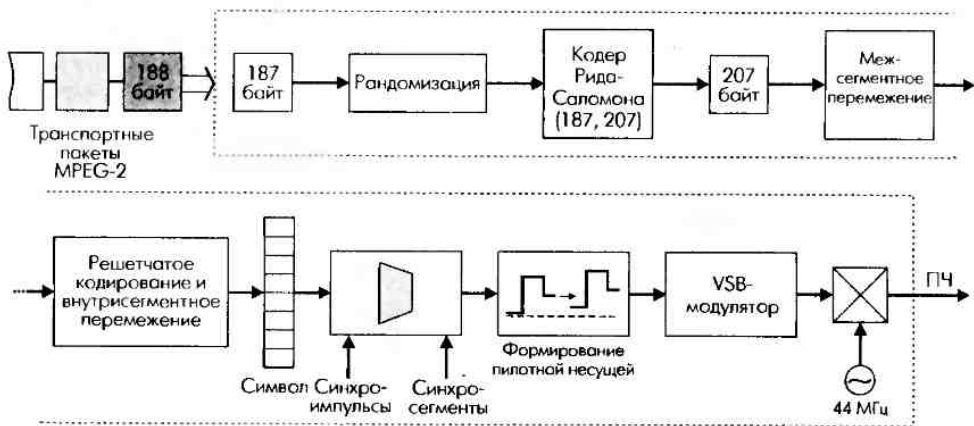


Рис. 2.51. Формирование сигнала в системе 8T-VSB

В итоге в 8T-VSB каждый байт превращается в 4 символа, 207 байт кодированного пакета — в 828 символов. Вместо синхробайта MPEG-2 используется синхросигнал (бросок амплитуды с уровня 5 до уровня -5) длительностью 0,37 мкс, что соответствует 4 символам. Получившиеся $828 + 4 = 832$ символа образуют сегмент. Каждые 312 сегментов данных и дополнительный синхросегмент объединяются в поле данных, два поля данных — в кадр (626 сегментов). Синхросегмент содержит ряд опорных последовательностей для настройки корректоров приемника, информацию о числе уровней модуляции и (только в 8T-VSB) 12 последних символов предшествующего сегмента.

Перед поступлением на вход кодера Рида–Соломона данные входного потока рандомизируют — превращают в псевдослучайные числа. Для этого каждый байт входного потока побитно складывается по модулю 2 с циклической псевдослучайной последовательностью. Генератором этой последовательности служит сдвиговый регистр из 16 триггеров, охваченный набором обратных связей. Во время синхронимпульса первого сегмента данных каждого поля в регистр загруз-

жается число $F180_{16}$. Восемь выходов регистра формируют поток байтов псевдослучайной последовательности. В результате сигнал становится практически независимым от изображения и шумоподобным, его спектральная плотность равномерно распределяется по полосе. Кроме повышения спектральной эффективности, улучшается и синхронизация передачи: поскольку уровни модуляции симметричны относительно 0, а амплитуды информационных символов достаточно случайны, средний уровень сигнала также близок к 0. На этом фоне легко детектировать периодически повторяющийся синхроимпульс и синхросегмент, даже при отношении сигнал/шум 0 дБ.

Поскольку средний уровень информационных сигналов близок к 0, амплитуда несущей при амплитудной модуляции также была бы нулевой. Но несущая в качестве пилот-сигнала необходима в данной системе для восстановления синхронизации в приемнике, поэтому вводят смещение — уровень каждого сигнала данных увеличивают на 1,25 единицы. Это соответствует появлению маломощного периодического сигнала несущей, добавляющего лишь 0,3 дБ к общей мощности сигнала. Данного уровня достаточно для детектирования несущей даже при отношении сигнал/шум 0 дБ.

Кодирование по алгоритму Рида–Соломона защищает информацию от кратковременных помех (до 10 ошибочных байт на 207-байтное слово). В случае длительных помех этот механизм сам по себе бессилен. Поэтому в VSB введено межсегментное перемежение — после кодера Рида–Соломона все байты одного пакета распределяются по 52 сегментам одного поля данных. Таким образом, даже если сегмент пропал, после деперемежения в каждом пакете окажется незначительное число поврежденных байтов, которые восстановятся декодером. Механизм перемежения используется и после решетчатого кодирования (в 8T-VSB), в результате которого последовательные символы оказываются отстоящими друг от друга на 12 символов — внутрисегментное перемежение. После внутрисегментного перемежения в сегменты вводят синхросигналы и добавляют синхропакеты полей, после чего смещают на 1,25 уровень сигналов данных для включения пилотной несущей. Столь сложным сигналом модулируется несущая на промежуточной частоте (ПЧ) 44 МГц, которая затем переносится непосредственно в полосу заданного телеканала.

К сожалению, имея лучшие теоретические показатели пороговых соотношений сигнал/шум и энергетической эффективности сигнала (на бит информации) [10], ATSC недостаточно надежен при многолучевом распространении сигналов, что неизбежно в условиях городской застройки. При этом отраженный сигнал, пришедший с задержкой, уже является помехой по отношению к основному. И если уровень отраженного сигнала менее чем на 15 дБ отличается от прямого, приемник теряет работоспособность — появляются помехи. То же самое относится и к приему сигналов от двух передатчиков. Поэтому заявленный выигрыш систем ATSC по отношению к DVB по показателю сигнал/шум в 4–6 дБ может проявиться разве что в лабораторных условиях. Практика же показывает, что в Нью-Йорке передатчик ATSC мощностью 350 кВт не обеспечивает 100% приема в радиусе 10 км, тогда как в Лондоне достаточно передатчика DVB-T мощностью 10 кВт для уверенного приема в радиусе 114 км [52]. Приходится усложнять антенные системы, что, очевидно, пользователей не радует.

Стандарт DVB

DVB, принятый в Европе, — это набор спецификаций, охватывающий кабельное DVB-C (cable), спутниковое DVB-S (satellite) и наземное DVB-T (terrestrial) вещание. Наиболее сложный алгоритм — в DVB-T, поскольку условия работы и требования к передаче при наземном вещании наиболее жесткие. Очень кратко остановимся на наиболее примечательных особенностях DVB.

Предварительная обработка пакетов в DVB в принципе аналогична 8-VSB, хотя механизмы реализации функций различны. На вход кодера поступают транспортные пакеты MPEG-2 по 188 байт (1 синхробайт (всегда 47_{16}) + 187 байтов данных) (рис. 2.52). Прежде всего они рандомизируются посредством сложения по модулю 2 с двоичной псевдослучайной последовательностью (генератор — 15-разрядный сдвиговый регистр). Генератор инициализируется через каждые восемь пакетов одним и тем же числом ($4B80_{16}$). Синхробайты не рандомизируют, каждый восьмой синхробайт инвертируют.

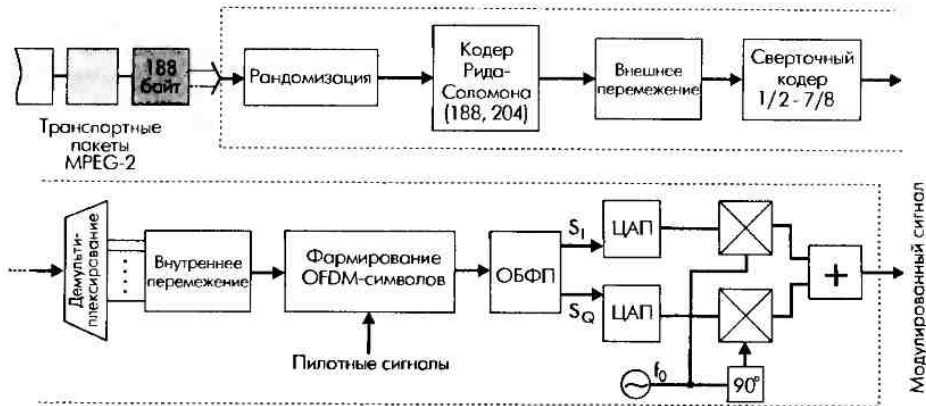


Рис. 2.52. Формирование сигнала в системе DVB-T

После рандомизации пакеты защищают кодом Рида-Соломона, в результате чего к 188 байтам добавляются 16 проверочных — возможно исправление до 8 ошибочных байт на кодированный 204-байтный пакет. Затем байты перемешиваются внутри кодированных пакетов, причем так, что синхробайты остаются на своих местах, — это внешнее перемежение. Затем следует внутреннее сверточное кодирование. Его реализует сдвиговый регистр из шести триггеров, превращающий каждый входной бит в два выходных (скорость кодера $1/2$). В DVB можно выбирать скорость сверточного кодирования ($1/2$, $2/3$, $3/4$, $5/6$, $7/8$), используя не оба элемента выходных пар, а лишь один. До сих пор функционально все было аналогично системе 8-VSB. Далее начинаются принципиальные различия, связанные с модуляцией радиосигнала.

В стандарте использована модуляция COFDM (Coded Orthogonal Frequency-Division Multiplexing) — вариант мультиплексирования посредством ортогональных несущих (OFDM) с предварительным кодированием сигнала. Модуляция OFDM подразумевает, что весь диапазон канала вещания (в Европе — 8 МГц) разбит на множество ортогональных поднесущих. Ортогональность означает, что

усредненное по времени произведение двух несущих равно нулю. Частоты поднесущих задаются как $f_n(t) = \cos 2\pi(f_0 + n/\tau)t$, где f_0 — нижняя частота диапазона, n — номер поднесущей ($n = 0..N - 1$; N — число поднесущих), τ — временной интервал передачи одного символа. Поток данных разбивается на N субпоток, несущая каждого из которых модулируется с гораздо меньшей скоростью. Разнос несущих по частоте равен $1/\tau$.

Поскольку в отдельном субканале скорость передачи невелика, перед каждым символом можно ввести защитный интервал — временной отрезок до $0,25\tau$, в течение которого транслируется фрагмент уже переданного символа (для сохранения ортогональности несущих). Основное назначение защитных интервалов — борьба с межсимвольными помехами, вызванными в том числе и переотражениями сигналов. Действительно, поскольку скорость символов мала, переотраженный сигнал в приемнике «накладывается» на прямо распространяющийся сигнал в интервале одного символа, а не следующего, попадая в защитный интервал.

Независимая (ортогональная) многочастотная передача с защитными интервалами позволяет успешно противостоять как мощным узкополосным помехам, так и переотраженным сигналам, причем методами цифровой обработки. В системе передачи с одной несущей VSB основное средство борьбы с аддитивными помехами — эквалайзер, однако при полном подавлении несущей сигнал восстановить уже невозможно. При многочастотной передаче «пропадут» только сигналы, попавшие в полосу помехи. Поскольку сигнал кодирован, информацию можно восстановить по данным из других субканалов.

В DVB в одном канале (при принятой в Европе ширине ТВ-канала 8 МГц) может быть до 8 тыс. несущих (учитывая требования алгоритмов обработки, $8 \times 1024 = 8192$ или 8К). Реально задействовано 1705 (режим 2К COFDM) или 6817 (режим 8К) несущих. Каждая несущая модулируется посредством 4-позиционной квадратурной фазовой манипуляции (QPSK) или 16- и 64-позиционной квадратурной амплитудной модуляции (QAM). Соответственно, на каждой несущей один модуляционный символ определяет от двух до шести бит.

Напомним, что при квадратурной модуляции выходной сигнал формируется сложением двух смещенных друг относительно друга на 90° гармонических колебаний на одной частоте f — синфазного $S_I(t) = A_S \cos \omega t$ и квадратурного $S_Q(t) = -A_Q \sin \omega t$ ($\omega = 2\pi f$). Их сумма — исходное колебание с фазовым сдвигом на угол $\varphi = \arctan(A_Q/A_S)$: $A_S \cos \omega t - A_Q \sin \omega t = A \cos(\omega t + \varphi)$; $A = \sqrt{A_S^2 + A_Q^2}$.

В соответствии с числом уровней модуляции исходный поток данных разбивается на n субпоток — по числу бит в модуляционном символе. Для QPSK таких субпоток два, для 16-QAM — четыре. Демультимплексирование происходит побитно — скажем, при модуляции 64-QAM ($n = 6$) первый бит попадает в первый субпоток, шестой — в шестой, седьмой — снова в первый и т. д. В DVB в каждом субпотоке биты переставляются по определенному правилу (своему для каждого субпотока) в пределах блока в 126 бит — внутреннее перемежение. Параллельные выходы устройств перемежения формируют модуляционный символ: 2-, 4- или 6-разрядный. На одной несущей OFDM передается один символ, поэтому в режиме 8К одновременно транслируется 48 групп по 126 символов —

всего $48 \times 126 = 6048$ информационных несущих (или 12 групп по 126 символов на 1512 несущих в режиме 2K). Одновременно передаваемые QAM-символы входят в OFDM-символ. Они распределяются по субканалам OFDM не последовательно, а опять-таки перемеживаются по определенному закону. Поэтому, если OFDM-символ пропадает, его данные можно восстановить, поскольку биты одного кодированного пакета оказываются распределенными по многим OFDM-символам.

Очевидно, что реализовать метод передачи OFDM «в лоб», т.е. использовать несколько тысяч генераторов модулированных поднесущих, весьма проблематично. А на приемной стороне это и вовсе неразрешимая задача. Однако современные методы цифровой обработки сигнала позволяют существенно упростить ее решение, используя отработанные алгоритмы прямого и обратного быстрого преобразования Фурье (БПФ и ОБПФ). Как это происходит?

Рассмотрим для примера векторную диаграмму модуляции 16-QAM (рис. 2.53). Каждая точка на ней соответствует четырем битам символа и определяет амплитуды синфазного (абсцисса) и квадратурного (ордината) колебаний. Складываясь, эти колебания задают соответствующий символу сигнал. Если применять математический аппарат комплексных чисел, ось квадратурных амплитуд будет соответствовать оси мнимых чисел (Im), ось синфазных амплитуд — оси действительных чисел (Re). Тогда любой символ можно представить как комплексное число $z = A_S + iA_Q$ или, по формуле Эйлера, $z = Ae^{i\varphi}$. В DVB используют не сами z , а их значения, нормированные на среднюю амплитуду суммарного колебания при выбранном виде модуляции. Это необходимо для усреднения амплитуд сигналов при различных режимах модуляции (очевидно, что при QPSK нормирующий множитель $c = 1/\sqrt{2}$, тогда как при 16-QAM $c = 1/\sqrt{10}$).

Запишем в комплексном виде модулированный символом $C_k(t)$ сигнал на k -й несущей (без ее переноса в диапазон вещания): $S_k(t) = C_k(t)e^{i2\pi kt/\tau}$. Учитывая, что сигнал изменяется дискретно, т.е. $t = nT$, где T — длительность дискретного интервала, n — номер отсчета, получим $S_k(nT) = C_k(nT)e^{i2\pi knT/\tau}$. Тогда общий сигнал OFDM-символа

$$S_n = \sum_{k=0}^{N-1} C_k(nT)e^{i2\pi knT/\tau}. \quad (2.3)$$

Математически это аналогично вычислению дискретных значений функции по дискретным значениям амплитуд (C_k) ее гармонических составляющих (с частотами k/τ). Данную процедуру описывает ОПФ:

$$X_n = \sum_{k=0}^{N-1} a_k e^{i2\pi kn/N}, \quad (2.4)$$

где N — число гармонических составляющих, в нашем случае — число несущих. Если длительность интервала дискретизации T выразить как τ/N , выражение (2.3) станет аналогичным ОПФ (2.4). Алгоритмы БПФ, как прямого, так

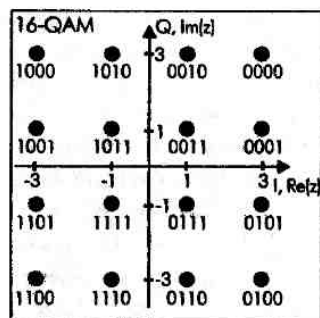


Рис. 2.53. Векторная диаграмма 16-QAM

и обратного, достаточно хорошо проработаны, в том числе и с точки зрения их аппаратной реализации. Они наиболее эффективны при N вида 2^m . Поэтому в 8K-COFDM число несущих условно принято равным $2^{13} = 8192$ — просто не все из них используются. Величина $1/T = N/\tau$ — это так называемая системная тактовая частота, для полосы канала в 8 МГц она равна 64/7 МГц. При переходе к другому частотному плану, например с полосой ТВ-канала 7 или 6 МГц, достаточно изменить системную тактовую частоту, сохраняя неизменной всю структуру обработки сигнала (а вместе с ней — основные функциональные устройства). Отметим, что системная тактовая частота одинакова в режимах 2К и 8К, т.е. от числа несущих скорость передачи напрямую не зависит, изменяется только надежность.

Таким образом, посредством ОБПФ из входного массива модуляционных символов численно формируется выходной OFDM-символ. Временной интервал его передачи складывается из собственно времени передачи символа τ и защитного интервала длительностью до $\tau/4$, в течение которого «повторно» передается часть символа (заключено в кавычки, поскольку защитный интервал следует перед информационным). Отметим, что кроме 6048 (в режиме 8К) информационных субканалов он включает еще пилотные сигналы, а также сведения о параметрах передачи — всего 6817 модулированных несущих. Пилотные сигналы — это фиксированные псевдослучайные последовательности с точно известными значениями фаз и амплитуд сигналов. Одна часть пилотных сигналов — непрерывные — передается на фиксированных несущих в каждом OFDM-символе, другая — распределенные — случайным образом (но равномерно) в произвольные моменты времени распределяется по спектральному диапазону передачи. Назначение пилотных сигналов — синхронизация и оценка параметров канала передачи.

Синтезировать OFDM-символы недостаточно, необходимо еще сформировать радиосигнал в заданной частотной области (с нижней частотой f_0). Перенос символа в необходимый диапазон — это его смещение на частоту f_0 , что в комплексной форме эквивалентно умножению на комплексное (в виде квадратурных слагаемых) представление несущей f_0 . При этом амплитуды перемножаются, а аргументы складываются. Выделяя действительную (синфазную) и мнимую (квадратурную) составляющие $S(n)$ и умножая их соответственно на $\cos(2\pi f_0 t)$ и $-\sin(2\pi f_0 t)$, после суммирования получим полный сигнал одного OFDM-символа.

Описанные механизмы позволяют очень гибко выбирать необходимый режим вещания, а также совмещать два потока пакетов MPEG-2 — с высокой и низкой скоростью. Возможную скорость определяют вид модуляции, скорость сверточного кодирования (СК), величина защитного интервала T_3 ($\tau/4$, $\tau/8$, $\tau/16$, $\tau/32$). Учитывая, что при 8K-OFDM $\tau = 896$ мкс, скорость изменяется в пределах от 4,98 Мбит/с (QPSK, СК = 1/2, $T_3 = \tau/4$) до 31,67 Мбит/с (64-QAM, СК = 7/8, $T_3 = \tau/32$).

Мы чрезвычайно схематично рассмотрели принцип передачи сигнала в DVB-T. Однако сигнал надо еще принять, демультимплексировать и декодировать, что сложнее, чем синтезировать его в передатчике. Для этого дополнительно к алгоритмам передатчика применяют корреляционные детекторы, декодеры с алгоритмом Витерби и т.д. Причем приемное устройство должно быть компактным

и недорогим иначе кто же его купит. Транспортные пакеты MPEG-2 также надо декодировать и сформировать ТВ-сигнал — цифровой или аналоговый, в зависимости от типа телевизора. Поэтому ЦТВ-приемник — это достаточно сложный программно-аппаратный комплекс, и только технологические достижения последних лет позволяют делать его недорогим при массовом выпуске.

На рынке уже появились DVB-T ресиверы в однокристальном исполнении, пример тому — ИС DVB-T-ресивера SQC 6100 компании Inficon. Данная ИС поддерживает практически все режимы DVB-T (2К/8К-COFDM; ТВ-канал 6, 7 и 8 МГц, все скорости сверточного кодирования и методы модуляции). Ее применение делает приемник DVB-T конструктивно простым (см. рис. 2.54). Аналоговый сигнал с антенны принимает ИС тюнера типа TUA 60xx. Он усиливает, фильтрует нужный ТВ-канал и преобразует его сигнал к первой ПЧ (36,125 МГц). После ПАВ-фильтра сигнал преобразуется смесителем (например, TDA 6190) ко второй ПЧ (7,225 МГц) и далее попадает на демодулятор SQC 6100. На выходе эта ИС формирует поток транспортных пакетов MPEG-2.

Первые наземные передачи ЦТВ и ТВЧ в США начались еще в 1997 году. По требованию Федеральной комиссии по связи (FCC) к маю 2002 года все коммерческие станции должны были перейти на передачи с использованием цифровых сигналов, для некоммерческих станций этот срок сдвигали на один год. К концу 2006 года все станции должны прекратить аналоговое вещание. Так изложено на бумаге. В действительности в США с цифровым телевидением ситуация не слишком радужная.



Рис. 2.54. DVB-T-приемник на базе ресивера SQC 6100

В отличие от происходившего в 60-е годы перехода на цветное телевидение, сигналы которого можно было принимать и на черно-белые приемники, сигнал ЦТВ аналоговый телевизор не воспринимает — необходим либо новый цифровой ТВ-приемник, либо цифровая принимающая приставка к аналоговому. Устройства эти, даже по американским меркам, не дешевые, например крупный магазин бытовой электроники в Нью-Йорке предлагает приемники ТВЧ Panasonic по цене от 1500 до 2800 долл. Около 650 долл. стоит ЦТВ-приставка Thomson. И хотя в стране уже более 170 телевизионных станций вещают в стандарте ATSC, продано всего лишь около 625 тыс. ЦТВ-приемников [156]. И это несмотря на то, что производством и поставками ЦТВ-приемников в США занимаются свыше 30 фирм, а парк аналоговых телевизоров в стране оценивают на уровне 300 млн [70]. Поэтому телекомпании как эфирного, так и кабельного вещания не спешат переходить на цифровой формат. А еще более усугубляет ситуацию конкуренция и видимые преимущества европейского стандарта DVB, который уже давно вышел за пределы Старого света.

Проблема усложняется и тем, что телезрители так и не получили обещанного высокого качества. Вернее, качество картинки высокое, однако надежность ее доставки оставляет желать лучшего. В этой связи в 1999 году представители

761 ТВ-станции США высказали озабоченность по поводу относительно низкого качества передаваемых ТВ-сигналов. Американская вещательная компания Sinclair Broadcasting Group — владелец наибольшего в США числа лицензий на ТВ-вещание направила в FCC доклад с призывом отказаться от принятой в ATSC модуляции 8-VSB и принять европейскую систему модуляции COFDM. Уже более 400 компаний объявили о поддержке этого доклада. Более того, Министерство обороны США высказало свою обеспокоенность низким (по сравнению с COFDM) качеством приема на портативные приемники 8-VSB-сигнала, что может оказаться существенным при объявлении общенациональной тревоги и в других чрезвычайных ситуациях. Поэтому Пентагон настаивает на внедрении COFDM в стандарт ATSC. Затянувшееся состояние неопределенности отнюдь не способствует развитию ЦТВ в США.

В Великобритании ситуация с цифровым телевизионным вещанием иная. Всего через два года после начала в ноябре 1998 года внедрения системы DVB 8 из каждых 10 домов уже находились в зоне действия наземного передатчика ЦТВ и могли принимать сигналы на антенну на крыше дома. Адаптеры — приставки для имеющихся телевизионных приемников — гораздо дешевле телевизоров стандарта ATSC, поэтому операторы платного телевидения предоставляют их телезрителям бесплатно. Около шести миллионов домов — порядка 20% общего их количества в Великобритании — принимают сигналы цифрового телевидения как от наземных станций платного телевидения (1 млн домов), так и от спутниковой системы BskyB (5 млн домов). Прекращение аналогового ТВ-вещания в Великобритании может начаться уже в 2006 году и закончиться к 2010 году. Однако, по мнению независимой телевизионной комиссии ИТС, хотя страна и является мировым лидером в данной области, даты перехода от аналогового телевизионного вещания к цифровому пока еще называть рано.

Переход к цифровому телевизионному вещанию в других странах мира

В развитии наземного цифрового телевизионного вещания остальная Европа отстает от Великобритании. Практически все европейские страны выражают намерение начать запуск DVB-систем в ближайшие два года. DVB-T вещание уже развертывается в Швеции, Испании, Финляндии и Дании, однако с меньшим успехом, поскольку адаптеры DVB-T там бесплатно не предоставляют. В Нидерландах внедрению DVB-T препятствует широко развитое кабельное телевидение.

За пределами Европы интерес к ЦТВ также велик. Канада и Мексика решили использовать систему ATSC, однако пока ожидают результатов сравнительных испытаний COFDM и 8-VSB. Аналогичная ситуация в Аргентине, Южной Корее и на Тайване. В Китае и Гонконге проводится сопоставление ATSC и DVB. Индия и Сингапур приняли DVB, но со звуковым стереосопровождением Dolby Digital surround system вместо MPEG. Австралия планирует использовать DVB, однако с изображениями высокой четкости и Dolby Digital. Сторонниками европейского стандарта выступают Нигерия и ЮАР.

Япония, подобно Европе, вначале пошла по пути развития аналогового телевидения со своим спутниковым 1125-строчным телевидением высокой четко-

сти Hi Vision. Теперь же в Японии разрабатывается цифровая система ISDB — во многом аналогичная DVB. В ISDB используется модифицированная форма COFDM с несколькими тысячами субканалов в 13 группах, несущих телевизионный сигнал с различной разрешающей способностью. В хороших условиях приема все группы принимаются и качество изображения наивысшее. Когда прием плохой, разрешающая способность снижается, однако качество изображения все же лучше, чем у обычного телевизионного вещания. Такой адаптивной деградации качества изображения у систем ATSC и DVB нет: там изображение либо заданного качества, либо вовсе отсутствует. В Японии выбрана и иная система звукового сопровождения — AAC (Advanced Audio Coding — перспективное кодирование аудиосигналов), предложенная германским Институтом им. Фраунгофера (одним из разработчиков стандарта MPEG), а также компаниями Dolby, Sony и AT&T. AAC позволяет вдвое уменьшить скорость передачи аудиоданных без потери качества. Япония призывает страны, которые не приняли систему ATSC или DVB, принять систему ISDB. Однако Сингапур после сравнительных испытаний выбрал европейский стандарт.

С системой цифрового телевидения жители России знакомы благодаря системе спутникового вещания «НТВ плюс» — с февраля 1999 года она транслирует цифровые программы в стандарте DVB-S. Известны проекты кабельного телевидения (в стандарте DVB-C), прежде всего компаний «Комкор» и «МТУ-Информ». Опытное вещание в стандарте DVB-T началось в Нижнем Новгороде 2 июня 2000 года. При этом использовались DVB-T-модулятор компании TANBERG Television (Великобритания), кодер MPEG-2 и мультиплексор НИИ Телевидения, а также передатчик «Онега 0.5Ц» АООТ «МАРТ» (Санкт-Петербург). Приемником служила DVB-приставка фирмы Nokia [2].

В целом развитие ЦТВ в России сдерживает не только отсутствие современной элементной базы [51]. Нет четкой государственной политики в области развития ЦТВ, если, конечно, не принимать за таковую ряд деклараций о намерениях, материально, увы, не подкрепленных. Иных преград для развития цифрового вещания в стране, где, по разным оценкам, от 60 до 85 млн телевизоров, не усматривается. Ведь нельзя же всерьез считать преградой отсутствие собственной элементной базы ЦТВ — практически весь мир использует не только импортную элементную базу, но и устройства на ее основе. Тем более что технологические проблемы вполне преодолимы — желающих производить по документации заказчика миллионные объемы ИС с самыми современными технологическими нормами сегодня достаточно. Наземное эфирное вещание, в отличие от кабельного и даже спутникового, — прерогатива именно государства, и без активных действий со стороны соответствующих государственных учреждений зоны опытного вещания такими и останутся.

Переход на ЦТВ в России — раньше или позже — неизбежен. Наша страна — пока открытый рынок для производителей аппаратуры ЦТВ. Кто его займет?

2.3.5 Широковещательные сети — цифровое радио

Задачи создания системы радиовещания, обеспечивающей в приемнике качество звучания, близкое к оригиналу, — заветная мечта не одного поколения радио-

инженеров. Сегодня она становится реальностью благодаря технологическому чуду конца прошлого века — возможности массового производства устройств с процессорами цифровой обработки аналоговых сигналов.

Серьезный инструмент повышения качества передаваемого аудиосигнала — цифровые форматы записи. Но возникает проблема. Как известно, качество записи на CD, рассматриваемое как эталонное, это 16-разрядная линейная оцифровка аналогового сигнала с частотой выборки 44,1 кГц на один канал. Современные алгоритмы компрессии способны сжать этот поток 1,5 Мбит/с более чем в 10 раз. Например, Musicam, ставший основой для знаменитого MPEG Audio Layer 2, — до 192 кбит/с. Чтобы транслировать такой поток, нужна достаточно широкая спектральная полоса. Где ее взять, и не одну, поскольку программ вещания достаточно много?

Единого подхода к решению подобной проблемы нет. На сегодня в мире сложились три основные технологии цифрового радио. Это европейская система Eureka-147, продвигаемая в США концепция IBOC и система низкочастотного (до 30 МГц) цифрового вещания DRM.

Система Eureka-147

История европейского стандарта цифрового вещания (Digital Audio Broadcasting — DAB) началась в 1987 году — с официальной даты рождения проекта Eureka-147. Разумеется, работы в области цифрового радио велись и до этого, например одной из составных частей нового стандарта стала разработка германского института IRT (Institut für Rundfunktechnik) 1981 года. В 1988 году в Женеве на Всемирной административной радиоконференции (WARC — World Administrative Radio Conference) уже демонстрировался первый образец оборудования для мобильного приема. В феврале 1995 года был опубликован европейский стандарт ETS 300 401, его вторая редакция вышла в марте 1997 года. Однако чаще его называют по имени проекта Eureka-147. Впоследствии, в 2000 и 2001 годах, спецификация ETS 300 401 уточнялась. Уже с 1995 года в европейских странах началось опытное цифровое вещание, с 1997 года — опытная коммерческая трансляция DAB-программ.

Что такое Eureka-147? В его основе [155] два механизма: кодирование аудиосигнала по методу Musicam (MPEG Audio Layer 2, используемое в известных стандартах MPEG 1 и 2) и разделение каналов посредством ортогональных несущих (OFDM), как в уже рассмотренной нами европейской системе цифрового телевидения DVB [44]. Общая схема передачи достаточно стандартна для современных коммуникационных технологий (рис. 2.55): аудиосигнал сжимается MPEG-кодеком, к нему добавляются информационные данные (название программы, курс валюты, прогноз погоды и т. п.) и служебная информация. Вся информация кодируется — сначала выравнивается (рандомизируется) амплитуда сигнала посредством умножения на заданную псевдослучайную последовательность, затем применяется сверточное защитное кодирование, после чего происходит временное перемежение информации. Кодированные таким образом каналы мультиплексируются в один основной сервисный канал. К пакетам этого канала

добавляется служебная информация — параметры мультиплексирования, информация о кодировании и т. д. Вместе с пакетами синхронизации вся информация образует кадр передачи, который транслируется посредством нескольких OFDM-символов (терминология стандарта ETS 300 401). OFDM-символ после квадратурной модуляции и обратного быстрого Фурье-преобразования представляет собой сигнал, которым можно непосредственно модулировать радионесущую. Полный сигнал эквивалентен низкоскоростной модуляции большого числа (от 1536 до 192) ортогональных несущих. Схема аналогична телевизионному цифровому стандарту DVB.

В полосе 1,536 МГц транслируется цифровой поток порядка 1,5 Мбит/с. Передача возможна в четырех режимах: моно, двухканальное моно, стерео и объединенном стерео (joint stereo). В последнем случае высокочастотные составляющие (выше 2 кГц) стереосигнала передаются в одном потоке, без разделения.

Кодек MPEG Audio Layer 2 предусматривает два режима кодирования аудиосигнала: с частотой выборки 24 и 48 кГц (от 8 до 160 и от 32 до 384 кбит/с). Соответственно, с частотой выборки исходный сигнал кодируется в диапазоне 11,3 и 20,3 кГц. Максимальная разрядность выборки 16 бит. В соответствии с алгоритмом весь частотный диапазон делится на 32 одинаковых субканала. В каждом из них разрядность оцифровки сигнала (импульсно-кодовой модуляции — ИКМ) определяется на основе психоакустической модели восприятия, в соответствии с которой восприимчивость уха к различным частотам неодинакова, следовательно, различна и разрядность кодирования. Кроме того, возможен и учет маскирования сигнала в одной полосе сигналом в другой, но это требует достаточно громоздких вычислений. Не углубляясь в механизм кодирования, отметим, что на выходе кодека формируется поток так называемых DAB-аудиокадров, включающих 1152 ИКМ-выборки аудиоинформации, связанные с ней данные и служебную информацию. Общая длительность кадра 24 мс при частоте сканирования 48 кГц. Значения выборок в каждом субканале нормируются с помощью так называемого фактора масштабирования (Scale Factor, SF), одинакового для каждых 12 выборок. Число бит в каждой выборке различно и зависит от номера субканала, выбранной скорости сжатия и частоты сканирования. Число бит в выборке субканала, фактор масштабирования и другая служебная информация передается в полях Bit allocation, SF, SCFS, а также в заголовке кадра (рис. 2.56).

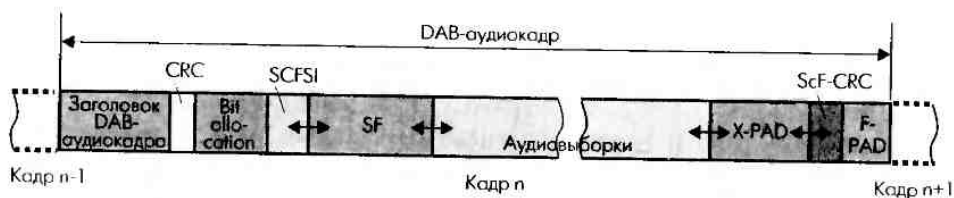


Рис. 2.56. Структура DAB-аудиокадра

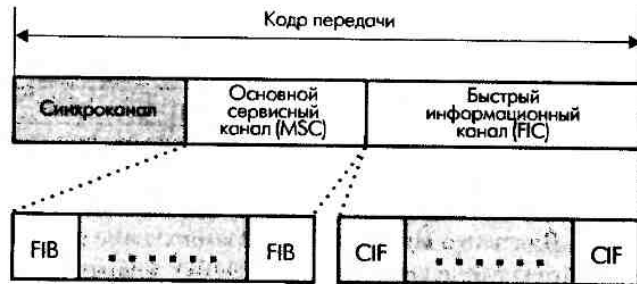
В кадрах MPEG Audio Layer 2 за полем аудиовыборок следует зарезервированное поле служебной информации. В DAB-аудиокадрах, структурно аналогичных MPEG, эти поля заняты двумя байтами «фиксированных данных, связанных с программой» (Fixed Programme Associated Data, F-PAD) и контрольной суммой

фактора масштабирования. Кроме того, по желанию пользователя может быть добавлено поле «дополнительных связанных с программой данных» (Extended PAD, E-PAD). Размер этого поля произволен — чем больше E-PAD, тем меньше поле аудиоданных.

Транспортный механизм стандарта ETS 300 401 предусматривает три основных информационных потока, формирующих кадр передачи (рис. 2.57):

- 1) основной сервисный канал Main Service Channel (MSC), служащий для передачи аудиоинформации и данных;
- 2) быстрый информационный канал Fast Information Channel (FIC), несущий служебную и вспомогательную информацию;
- 3) канал синхронизации, используемый при демодуляции (кадровая синхронизация, контроль частоты, оценка состояния канала, идентификация передатчика).

Рис. 2.57. Структура кадра передачи



Быстрый информационный канал FIC состоит из так называемых быстрых информационных блоков (Fast Information Block, FIB). Каждый FIB включает 240 бит полезной информации и 16-битную контрольную сумму CRC, всего 256 бит. Поле полезной информации, в свою очередь, может быть разбито на сегменты (Fast Information Groups, FIGs) произвольной длины. Основное назначение FIC — передача многочисленной служебной информации, а также поддержка различных информационных сервисов (индикация времени, радиотекст, меню программ, пэйджинг и т. п.). По этому же каналу передается информация о мультиплексировании различных программ и сервисов в кадре передачи.

Основной сервисный канал MSC состоит из Common Interleaved Frames (CIF) — общих вложенных кадров, по 55 296 бит каждый. Минимальный адресуемый элемент CIF — Capacity Unit (CU) 64 бит. Всего в CIF 864 CU. CIF подразделяется на субканалы. Каждый из них включает целое число последовательных CU. Субканалу соответствует определенная радиопрограмма (сервис). Очевидно, что допустимое число программ и дополнительных сервисов (субканалов) в кадре передачи зависит от скорости кодирования аудиосигнала. Практика показала, что в одной полосе возможно транслировать до шести программ с качеством аудиосигнала на уровне CD.

По основному сервисному каналу передаются не только аудиоданные, но и информация общего назначения. Стандартом предусмотрено два режима передачи

данных — потоковый и пакетный. Скорость данных в первом режиме кратна 8 кбит/с, во втором — зависит от размера и частоты следования пакетов. Пакетный режим позволяет в одном субканале передавать несколько различных сервисов, поскольку размер пакета невелик — от 24 до 96 байт.

Стандарт Eureka-147 задумывался как максимально универсальный — для наземного, кабельного и спутникового вещания, для одночастотных глобальных сетей и локальных зон. Поэтому в нем предусмотрено четыре режима передачи, каждый из которых оптимален для определенного применения. Назначение этих режимов рассмотрим ниже, сейчас лишь отметим, что структура и длительность кадра передачи в каждом из них различны (табл. 2.9).

Таблица 2.9. Структура кадра в зависимости от режима передачи

Режим передачи	Длительность кадра, мс	Число FIB в кадре	Число CIF в кадре
I	96	12	4
II	24	3	1
III	24	4	1
IV	48	6	2

В MSC все данные кодируются, причем независимо для каждого подканала. Прежде всего информация может быть защищена от несанкционированного доступа. Для этого используется перемножение информации на некоторую псевдослучайную последовательность (ПСП), ключом является 8-байтное контрольное слово, единожды заданное или периодически изменяемое и передаваемое в приемники. Система контроля доступа может и не использоваться. Однако это — инструмент авторизации программ, взимания абонентской платы за их прослушивание, а при желании — за каждый сервис в отдельности.

Далее все данные кодируются в целях энергетического сглаживания сигнала, чтобы его спектр был как можно более ровным, шумоподобным. Для этого используется умножение на ПСП, которую формирует генератор с задающим полиномом вида $P(x) = x^9 + x^5 + 1$ (рис. 2.58). Инициализирующее слово — $1FF_{16}$.

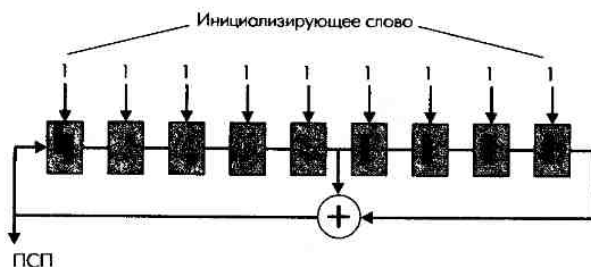
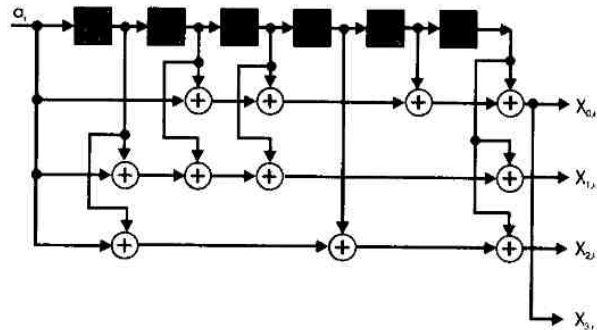


Рис. 2.58. Генератор ПСП для рандомизации спектра сигнала

После выравнивающего кодирования начинается сверточное избыточное кодирование, при котором объем информации существенно (на 25–300%) возрастает, а вместе с ним — и степень защиты от ошибок передачи. Схема сверточного кодера приведена на рис. 2.59. Каждому входному биту соответствуют четыре выходных. Однако используют не обязательно все четыре, поэтому

скорость кодирования может варьироваться от $8/9$ до $8/32$ (число входных/выходных бит). Существенно, что различные информационные поля кодируются с разной скоростью. Кроме того, скорости кодирования аудиоинформации в MSC и данных в FIC различны. В итоге наиболее защищенной оказывается служебная управляющая информация, наименее — поля аудиовыборок, что естественно: потери управляющей информации могут носить фатальный характер. Скорость кодирования в FIC — $1/3$, т. е. вместо исходных 256 бит FIB образуются $3 \times 256 = 768$ бит.

Рис. 2.59. Сверточный кодер



После сверточного кодирования данные в MSC — и только в нем — подвергаются временному перемеживанию (перемешиванию). В этом канале вся информация структурирована в виде логических кадров — пакетов информации, связанных с исходными блоками данных длительностью 24 мс (DAB-аудиокадры либо блоки данных). Число бит в логическом кадре зависит от стадии и скорости кодирования. Информация из каждого 16 логических кадров перемешивается по определенному закону и распределяется по 16 последовательным пакетам данных. Благодаря этому, если из-за длительной помехи какой-либо пакет пропадет, исчезнет лишь часть информации каждого из 16 логических кадров, которую можно будет восстановить при декодировании.

В MSC все процедуры кодирования выполняются независимо для каждого аудиопотока или информационного сервиса. И только после временного перемеживания пакеты мультиплексируются в CIF. Каждому потоку в CIF выделяется субканал, занимающий целое число 64-битных CU. Информация о схеме мультиплексирования (начальный адрес и размер субканала, скорость кодирования) передается в FIC. После формирования CIF в соответствии с выбранным режимом передачи (см. табл. 2.9) мультиплексор формирует кадр передачи (рис. 2.60).

Информация кадра передачи транслируется в виде последовательности OFDM-символов. Первые два OFDM-символа кадра принадлежат каналу синхронизации. Это так называемый NULL-символ (уровень сигнала равен 0) и символ, несущий информацию о фазе дифференциальной QPSK-модуляции в следующем OFDM-символе. Информация из FIB и CIF группируется в виде битовой последовательности и разбивается на OFDM-символы. Каждому символу соответствует число бит, вдвое большее, чем число несущих в выбранном режиме передачи (табл. 2.10). Например, для первого режима передачи одному OFDM-символу соответствуют $2 \times 1536 = 3072$ бита. Поскольку размер кодированного

FIB — 768 бит, в этом режиме передачи одному OFDM-символу соответствует 4 кодированных FIB, или 3 OFDM-символа на все 12 пакетов быстрого информационного канала. В основном сервисном канале размер пакета FIC — 55 296 бит, или 18 OFDM-символов на пакет в первом режиме передачи (всего 72 пакета на 4 CIE-кадра передачи).

Таблица 2.10. Параметры режимов передачи Eutrema-147

Режим передачи	I	II	III	IV
Длительность кадра, мс	96	24	24	48
Длительность Null-символа, мкс	1297	324	168	648
Длительность защитного интервала, мкс	246	62	31	123
Номинальное максимальное расстояние между передатчиками, км	96	24	12	48
Номинальный частотный диапазон, МГц	до 375	до 1500	до 3000	до 1500
Длительность информационной части символа, мкс	1000	250	125	500
Общая длительность символа, мкс	1246	312	156	623
Число ортогональных несущих	1536	384	192	768

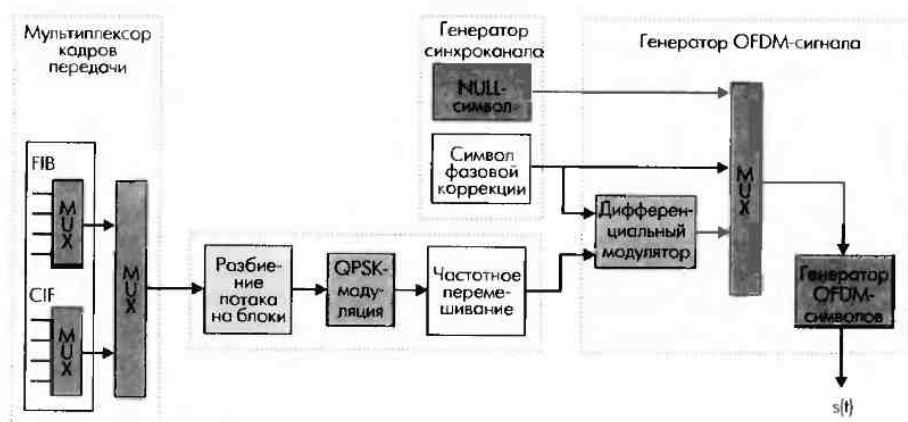


Рис. 2.60. Функциональная схема формирования радиосигнала

Далее последовательность OFDM-символов кадра, за исключением первых двух из синхронизации, подвергается квадратурной модуляции QPSK. Каждым двум битам двоичного OFDM-символа ставится в соответствие QPSK-символ, который в комплексном виде может быть записан как $q_{l,n} = \frac{\sqrt{2}}{2}[(1 - 2p_{l,n}) + i \times (1 - 2p_{l,n+k})]$, где $n = 0, 1, \dots, K - 1$ — номер несущей, $l = 2, \dots, L - 1$ — текущий номер OFDM-символа (всего L символов в кадре), $p_{l,n}$ — n -й бит l -го OFDM-символа, K — число ортогональных несущих в заданном режиме переда-

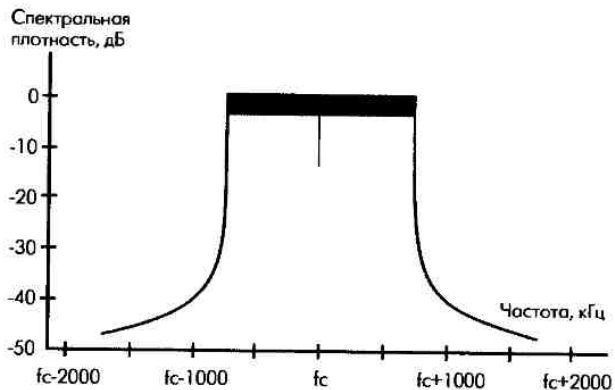
чи. Иными словами, два бита раскладываются на квадратурные составляющие одной несущей.

Далее происходит частотное перемешивание: последовательные QPSK-символы распределяются по K несущим по определенному закону. В результате помеха в полосе нескольких соседних несущих повредит лишь малую часть информации из разных информационных блоков. Итоговая последовательность QPSK-символов попадает в дифференциальный QPSK-модулятор со сдвигом фазы на $\pi/4$, поэтому и необходим второй символ синхроканала с информацией о фазе. К модулированной последовательности добавляются два символа из синхроканала, и последовательность поступает в генератор OFDM-сигнала, в котором формируется аналоговый сигнал. Отметим, что передаваемые OFDM-символы разделены защитным временным интервалом, что избавляет от помех, связанных с многолучевым распространением (переотражениями), неизбежным в городских условиях, — переотраженный сигнал попадает в защитный интервал и не интерферирует со следующим OFDM-символом.

К основному DAB-сигналу может быть добавлен так называемый сигнал-идентификатор передатчика, размещаемый в NULL-символах некоторых кадров передачи. Его основное назначение — передавать информацию о географическом положении передатчика.

Спектр DAB-сигнала в любом режиме передачи занимает полосу порядка 1,5 МГц (рис. 2.61). Параметры режимов передачи выбраны так, что первый режим оптимален для одночастотных наземных сетей вещания в УКВ-диапазоне, режимы 2 и 4 ориентированы для наземной трансляции в УКВ-, а также в L-диапазоне (1452–1492 МГц) в локальных и одночастотных глобальных сетях. Режим 3 лучше всего подходит для кабельного и спутникового вещания, а также для гибридных спутниково-наземных сетей.

Рис. 2.61. Теоретический спектр DAB-радиосигнала в первом режиме передачи



Мы столь подробно остановились на Eureka-147, поскольку это европейский, следовательно, наиболее близкий для России стандарт. Кроме того, это единственный на сегодня стандарт цифрового радиовещания, действительно обеспечивающий CD-качество программ. На этом можно было бы поставить точку, если бы не важнейший вопрос: в каком диапазоне организовывать вещание? Ведь

стандарт рассчитан на диапазоны в пределах от 30 МГц до 3 ГГц. И в России, и в США в этой части радиоспектра свободных полос нет — они заняты действующими в полосе 87,5–108 МГц FM-радиостанциями или зарезервированы под военные и им подобные нужды. Быстрого отказа от FM-вещания не предвидится, поскольку цены DAB-приемников — несколько сот долларов, а в США — сотни коммерческих FM-радиостанций. Поэтому Федеральная комиссия связи США FCC не дала добро на внедрение Eureka-147. Там пошли своим путем, развивая систему iDAB на базе технологии IBOC (In Band On Channel).

Технология IBOC

Основная идея IBOC — сделать переход от аналогового вещания к цифровому плавным, без выделения дополнительных каналов. Рассматриваются два диапазона: занятые сегодня под вещание с амплитудной модуляцией (AM) длинные, средние и короткие волны и УКВ-диапазон с частотной модуляцией (FM) 87,5–108 МГц. В каждом из них предусмотрено два режима вещания: гибридный и полностью цифровой. В первом случае происходит одновременное аналоговое и цифровое вещание, во втором — только цифровое.

В FM-режиме основная полоса каналов занимает около 240 кГц, а две боковые — еще по 120 кГц. По правилам FCC, уровень сигнала в боковых полосах должен быть на 25 дБ ниже, чем в основной (рис. 2.62). В гибридном режиме сигналы цифрового вещания передают в боковых полосах, в цифровом — в основной и боковых. Теоретически возможная скорость цифрового потока в боковых полосах 64 кбит/с, в основной полосе — 96–128 кбит/с.

В AM-вещании модулируется несущая с частотой 9/10 кГц. Реальная ширина спектральной полосы основного сигнала — порядка 20 кГц, еще по 10 кГц занимают боковые полосы (рис. 2.63). В гибридном режиме полоса цифрового потока перекрывает полосу аналогового вещания, но с меньшим уровнем сигнала. В цифровом режиме сигнал занимает всю полосу в соответствии с разрешенным шаблоном. Скорость передачи в гибридном режиме от 20 до 48 кбит/с, в полностью цифровом — до 64 кбит/с.

Чтобы в столь узких полосах, особенно в AM-диапазонах, достичь ощутимого качества вещания, необходим совершенный аудиокодек. Система IBOC основывалась на технологии кодирования PAC (Perceptual Audio Coding) компании Lucent Technologies. Одно из его важнейших свойств в том, что информация разбивается на отдельные потоки: основной и дополнительные. Если ширина канала не позволяет передавать всю информацию, транслируется только основной поток с некоторой потерей качества аудиосигнала. Если появляется возможность передавать дополнительные потоки, качество повышается. При PAC-кодировании при скоростях потока 128–96 кбит/с качество декодированного сигнала мало отличается от оригинала — стереосигнала с CD-диска (16 бит, частота выборки 44,1 кГц). Но такие скорости возможны лишь при полностью цифровом FM-режиме. В AM-режимах при скорости потока 48 кбит/с качество цифрового вещания примерно соответствует хорошему приему аналогового FM-стерео.

Развитие технологии цифрового вещания в США сегодня сосредоточено в образованной в августе 2000 года корпорации iBiquity Digital. Эта компания ро-

делась в результате совместной двухлетней работы корпораций USADR (USA Digital Radio) и Lucent Digital Radio. USADR возникла в 1991 году как альянс компаний CBS, Gannett и Westinghouse Electric. С 1998 года она преобразована в независимую компанию. Именно ей принадлежат ключевые патенты на технологию IBOC. Lucent Digital Radio появилась также в 1998 году как подразделение Lucent Technologies в союзе с фирмой Pequot Capital Management. Компания развивала собственную версию DAB на основе технологии аудиокодирования AAC.

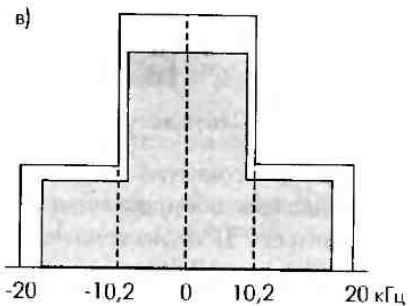
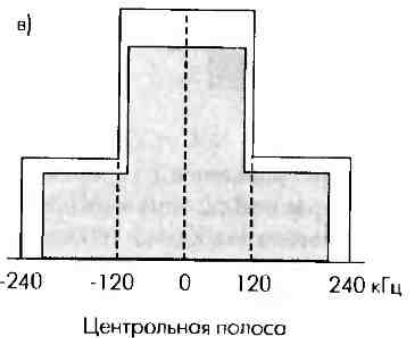
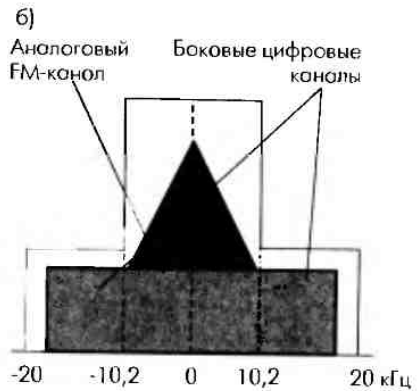
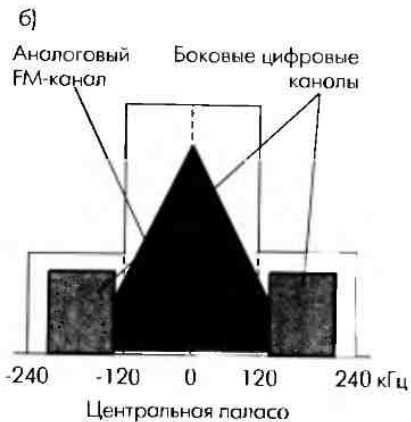
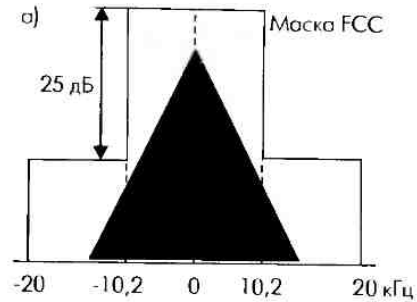
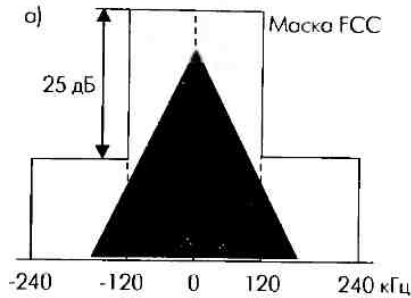


Рис. 2.62. Спектр сигналов IBOC в диапазоне FM-вещания: а — аналоговый, б — гибридный, в — цифровой режимы

Рис. 2.63. Спектр сигналов IBOC в диапазоне АМ-вещания: а — аналоговый, б — гибридный, в — цифровой режимы

Существенная особенность технологии цифрового вещания в США — возможность передавать одну и ту же программу в цифровом и аналоговом виде с некоторым временным сдвигом. Когда цифровой сигнал пропадает, приемник незаметно для слушателя переходит на воспроизведение аналогового сигнала. Это позволяет успешно бороться с так называемым клифф-эффектом, проявляющимся в цифровых вещательных системах в виде периодов неразборчивости сигнала. По мнению специалистов компании iBiquity, переход на полностью цифровое вещание нецелесообразен, пока число IBOS-приемников не составит по меньшей мере 85%.

Система вещания IBOS строится на технологии модуляции с использованием ортогональных несущих COFDM. Структура передатчика приведена на рис. 2.61. Аудиосигнал параллельно транслируется в цифровом и аналоговом виде. В цифровом тракте происходит аудиокодирование, защитное кодирование и перемежение, после чего цифровая последовательность поступает на OFDM-модулятор. Аналоговый сигнал с задержкой поступает в стандартный аналоговый тракт и после обработки смешивается с цифровым сигналом OFDM-модулятора. Суммарный сигнал после фазово-амплитудной модуляции попадает в стандартный передающий модуль.

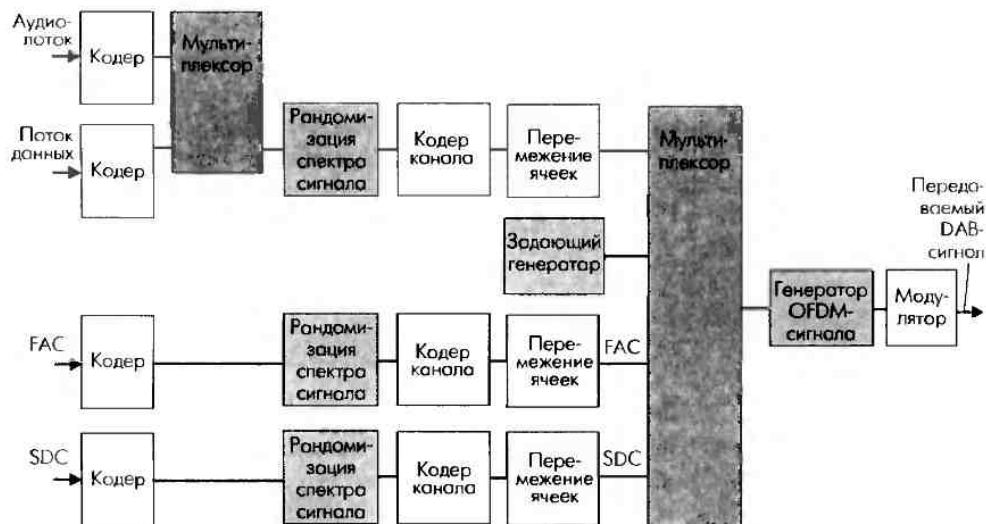


Рис. 2.61. IBOS-передатчик

Основное достоинство американской системы по сравнению с европейской — это более дешевое оборудование, прежде всего — приемники. Еще несколько лет назад стоимость DAB-ресивера для Eureka-147 составляла около 1000 долл., в то время как для IBOS-приемников назывались цифры порядка 100 долл. Но ситуация стремительно меняется, и хотя цены на приемники Eureka-147 все еще достаточно высоки, однако они уже достигли уровня 300–400 долл., и тенденция к их снижению продолжится.

Кроме наземных систем вещания, важную роль в США отводят спутниковому цифровому радиовещанию. Здесь реально конкурируют две системы: компаний

Sirius Satellite Radio и XM Satellite Radio. С технической точки зрения эти спутниковые системы во многом уникальны. Поскольку они предназначены в основном для работы с автомобильным радио и ориентированы на массовый потребительский рынок, их приемники должны быть дешевле, чем любой ранее выпущенный ресивер спутниковых сигналов. Приемные антенны — обязательно ненаправленные, с низким усилением, поскольку дорогостоящие системы слежения антенны за спутником неприемлемы.

В системе XM оба спутника находятся на обычной геостационарной орбите в точках стояния 85° и 115° западной долготы. Под именами Rock и Roll они были выведены на орбиту соответственно в марте и мае 2001 года. По заявлению представителей компании, это «наиболее мощные спутники в индустрии развлечений», что помогает компенсировать относительно малый коэффициент усиления приемной антенны на автомобиле.

Оба спутника транслируют идентичные сигналы в разных частотных диапазонах. Их пространственное положение таково, что при идеальных условиях приемник будет «видеть» и принимать сигналы от обоих спутников. Поэтому затенение одного из них не приведет к пропаданию сигнала. Кроме того, сигналы транслируются не синхронно, а с некоторой временной задержкой (до 4–5 с), поэтому, если оба спутника на короткое время исчезнут из поля зрения приемника, передача все равно не прервется.

В системе компании Sirius Satellite Radio задействовано три спутника на сильно вытянутых эллиптических орбитах. Период обращения спутников — примерно 16 часов, поэтому в зоне видимости с территории США всегда находятся два из них. В среднем угол возвышения спутников составляет 60° , что больше угла возвышения (45°) спутников на геостационарной орбите системы XM. В результате в системе Sirius вероятность экранирования спутника наземным препятствием ниже, а само по себе постоянное перемещение спутников — не проблема, поскольку приемные антенны практически всенаправлены.

Поскольку для передачи отведены лишь два частотных канала, только два из трех спутников могут одновременно передавать сигналы, иначе два спутника работали бы в одном частотном диапазоне и создавали взаимные помехи. Поэтому у опускающегося за горизонт спутника ретранслятор выключается, у поднимающегося — включается. При этом возможен кратковременный (порядка миллисекунд) перерыв приема от двух спутников, однако третий — наиболее высокий спутник — продолжает вещание, поэтому подобный перерыв практически незаметен.

Как в системе XM, так и в системе Sirius предусмотрена сеть наземных ретрансляторов для обслуживания таких недоступных для спутников участков, как туннели или «городские каньоны». В отличие от применяемой при вещании со спутников четырехпозиционной квадратурной модуляции QPSK с временным разделением каналов (TDM), наземные ретрансляторы используют COFDM/TDM-модуляцию. В системе XM со спутниками на геостационарной орбите ретрансляторы (всего 1500) принимают сигналы с помощью неподвижных направленных антенн с высоким коэффициентом усиления. Это необходимо, поскольку переизлучаемый ретранслятором сигнал очень близок по частоте к принимаемому со спутника, но гораздо более мощный. Поэтому он не должен попадать на при-

емную антенну. В системе Sirius для узконаправленной антенны понадобилась бы система слежения. Вместо этого выбран более дешевый метод передачи сигналов на 105 наземных ретрансляторов от геостационарных связных спутников, которые работают в другом частотном диапазоне (Ku, 12–14 ГГц).

Вещание на приемники ведется в диапазоне около 2,3 ГГц, общая скорость информационного потока (до кодирования) — 4 Мбит/с в XM и 4,4 Мбит/с — в Sirius, что соответствует примерно 50 музыкальным программам (по 64 кбит/с) и 50 новостным. Аудиосигнал кодируется PAC-алгоритмом, для защиты от ошибок используется кодер Рида-Соломона и сверточное кодирование со скоростью 1/2.

Многие известные компании, например Pioneer, Alpine, Sony, Blaupunkt, Clarion, Delphi-Delco Electronics System, производят приемники для спутниковой системы XM. Причем цена такого приемника на 150 долл. выше обычного AM-FM. Для системы Sirius приемники производят такие компании, как Kenwood, Clarion, Alpine, Jensen, Panasonic и Visteon. Компания Sirius Satellite Radio имеет соглашения на установку трехдиапазонных (AM/ЧМ/SAT) приемников в автомобили Ford, Chrysler, BMW, Mercedes, Mazda, Jaguar и Volvo, а также в тяжелые грузовые автомобили Freightliner и Sterling.

Всемирное цифровое радио (DRM)

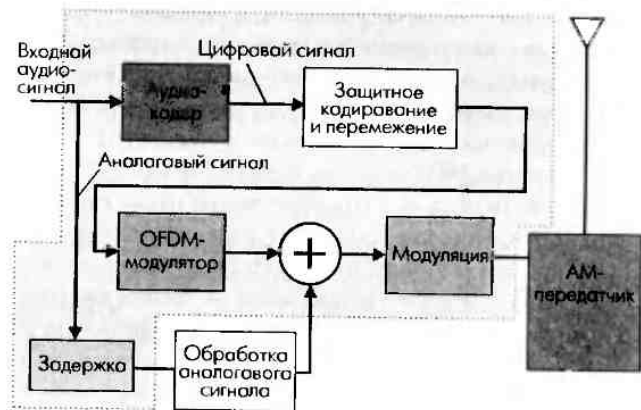
DRM (Digital Radio Mondiale, всемирное цифровое радио) — это действительно международный консорциум, объединяющий свыше 70 известных исследовательских, производственных и вещательных компаний и организаций всего мира. Его цель — создание стандарта цифрового радиовещания в полосе ниже 30 МГц. При этом преследуются такие цели, как значительное улучшение качества звука, снижение помех и сохранение зоны обслуживания при значительном — до шести раз — уменьшении мощности передатчика. Поскольку ширина полосы цифрового вещательного сигнала в DRM эквивалентна полосе аналогового, для перехода на цифровое вещание дополнительного спектрального диапазона не требуется. Образованный в 1996 году, этот консорциум уже стал реальным игроком в мире цифрового вещания — многие ведущие производители уже демонстрируют опытные образцы DRM-оборудования. Спецификация DRM признана как ITU (Рекомендация ITU-R BS.1514), так и европейским институтом стандартов ETSI.

Система DRM также использует OFDM-модуляцию (порядка 200 несущих) с предварительной 16- и 64-позиционной квадратурной амплитудной модуляцией (16- и 64-QAM), защитное сверточное кодирование и временное перемежение информации. Для аудиокодирования принят алгоритм AAC (Advanced Audio Coding) — низкоскоростной алгоритм, применяемый в группе стандартов MPEG-4 (изначально создавался для аудиокодирования в MPEG-2). Кроме того, для кодирования речи предусмотрен низкоскоростной алгоритм CELP. Механизм AAC дополнен алгоритмом Spectral Band Replication (SBR) компании Coding Technologies (Германия), позволяющим существенно улучшить качество принимаемого аудиосигнала без существенного увеличения скорости потока. Ограничение скорости цифрового потока при кодировании приводит к сужению спектра аудиосигнала — пропадают верхние частоты. Алгоритм SBR позволяет восстанавливать высокочастотную составляющую сигнала по переданной низ-

кочастотной, крайне незначительно увеличивая необходимую скорость потока аудиоданных. Так, если после сжатия кодером ACC верхняя частота передаваемого аудиосигнала 6 кГц, то SBR поднимает ее до 15 кГц.

В DRM предусмотрено три информационных канала (рис. 2.65): Main Service Channel (MSC), Fast Access Channel (FAC) и Service Description Channel (SDC). По MSC передаются оцифрованные аудиосигналы и информационные данные (всего до четырех различных потоков), через FAC — оперативная служебная информация, SDC используется для передачи относительно редко обновляемой служебной информации. Передача через MSC и FAC происходит кадрами длительностью 0,4 с, через SDC — кадрами по 1,2 с [154].

Рис. 2.65. Передающая система DRM



Таким образом, сегодня существует несколько систем цифрового радио, которые можно рассматривать как стандарты, что и подтверждает в своих недавних рекомендациях ITU. Причем если Eureka-147 ориентирована на широкополосные одночастотные трансляционные сети и в ряде случаев требует пересмотра сложившегося распределения частот, то системы DRM и IBOC изначально предназначены для работы в уже существующих диапазонах. Недаром наряду с развитием систем Eureka-147 многие европейские производители и радиовещатели (например, BBC) уделяют серьезное внимание DRM. Действительно, традиционное АМ-вещание в диапазоне средних и коротких волн умирает, поскольку никого не устраивает по качеству. Цифровое же вещание позволит использовать достоинства распространения волн в этих диапазонах с качеством, конечно, не СД, но уж по крайней мере традиционного FM-вещания. Сколько при этом будут стоить приемники — вопрос.

WACR-92 отвела полосу частот 1452–1492 МГц для спутникового цифрового радиовещания (кроме США, России, Белоруссии и Украины). В этой же полосе возможно создание и наземных сетей. Но до 2007 года эти частоты не могут быть использованы на первичной основе. Система же Eureka-147 сегодня развивается весьма активно и не только в Европе, но и в Австралии, Канаде и Израиле. Аудитория DAB-слушателей во всем мире оценивается свыше 284 млн человек. Поскольку цифровое радиовещание ориентировано прежде всего на автотранспорт, системы ретрансляции строят вдоль крупнейших автотрасс. Сегодня под

непрерывный аккомпанемент DAB-станций можно проехать от Женевы до Мюнхена (468 км), в Норвегии — из Порсгунда в Тронхейм (800 км), из Лондона в Ньюкасл (459 км) и т. д.

Кроме собственно передачи звуковых программ, DAB Eureka-147 может обеспечить и различные информационные сервисы, в том числе низкоскоростное мобильное видео (MPEG-2), трансляцию различных картинок (вспомним, что явилось толчком и экономическим фундаментом развития Интернета), различную текстовую новостную и справочную информацию. Все это делает цифровое вещание мощным конкурентом третьего поколения (3G) сотовой связи в области информационных услуг и мобильного видео. Хотя DAB не интерактивна, но разнообразие сервисов DAB в известной мере компенсирует этот недостаток. При этом лицензии на DAB-вещание, например, в Великобритании (порядка 50 тыс. фунтов) несопоставимо дешевле миллиардных 3G-лицензий сотовой связи.

Наступление эры цифрового радио неизбежно и в нашей стране. «Концепция внедрения наземного цифрового звукового вещания в России в части использования радиочастотного спектра» существует с 1998 года. В феврале 2002 года Министерство РФ по делам печати, телерадиовещания и средств массовых коммуникаций передало в Правительство проект постановления «О мерах по обеспечению внедрения цифрового телевизионного и звукового вещания в Российской Федерации». Он предусматривал разработку в 2002–2003 годах программ внедрения телевизионного и звукового вещания, частотного плана для этого вещания по всей территории РФ, положения о конкурсе на право использования радиочастотного спектра для наземного телерадиовещания, а также ряда государственных стандартов и создание соответствующей аппаратуры. Пока все осталось на бумаге.

Видимо, в России, ориентирующейся на систему Eureka-147, одна из наиболее сложных проблем — поиск частотного ресурса. В качестве наиболее вероятного указывают на так называемый III вещательный диапазон 174–230 МГц. Однако все может упереться в основную проблему: где взять финансирование на разработку нормативных документов, разработку оборудования и развертывание сетей.

2.4 Спутниковые сети

2.4.1 Исторический очерк развития спутниковой связи

Артур Чарльз Кларк в статье «Extra-terrestrial Relays», опубликованной в 1945 году в журнале *Wireless World*, пророчески писал: «Искусственный спутник земли на определенном расстоянии от поверхности будет совершать один оборот каждые 24 часа. То есть постоянно оставаться над одной точкой земного шара. Из-за чего он будет в пределах прямой видимости почти половины поверхности земли. Три ретрансляционных станции, расположенные на этой орбите в пределах 120 градусов, позволят охватить всю планету телевидением и радиосвязью».

Через 15 лет, как только появилась техническая возможность вывода искусственных спутников Земли, начались и работы в области создания спутниковых систем связи (ССС). Первым спутником связи стал запущенный 12 августа 1960

года американский космический аппарат (КА) Echo-1. Это был пассивный ретранслятор — надувной шар диаметром порядка 30 м (100 футов) с металлизированной оболочкой на орбите немногим выше 1,5 тыс. км. Он служил экспериментальным целям, в дальнейшем направление пассивных ретрансляторов в силу неэффективности развития не получило.

Активный ретранслятор впервые выведен на орбиту 10 июля 1962 года — им стал спутник Telstar I. Орбита этого КА оказалась в районе так называемого первого радиационного пояса Ван-Аллена. В результате все его электронное оборудование вышло из строя за несколько недель. Запущенный на следующий год Telstar II был более радиационностойким и проработал гораздо дольше. Он позволял передавать один телевизионный канал или 60 дуплексных телефонных каналов. Через Telstar II была организована первая успешная трансатлантическая телепередача. Однако орбита этого КА была низкой, а мощность передатчика не превышала 2 Вт. Все это требовало больших наземных антенн со сложной системой сопровождения спутников, а также делало невозможным продолжительные сеансы связи.

Но уже в 1963 году начались запуски КА серии Syncom на геостационарные орбиты. Первые две попытки были неудачны (14 февраля и 26 июля, в последнем случае орбита была почти геосинхронной, период обращения составлял 24,2 часа, но с наклоном около 30°). Но 19 августа 1964 года Syncom-3 занял свое место на геостационарной орбите [99]. Связь через него обеспечивали две наземные станции с 9-метровыми антеннами — в городе Лейкхерсте (США, шт. Нью-Джерси) и на судне «Кингспорт», стоявшем в Гвинейском заливе близ нигерийского города Лагос [42]. Syncom-3 использовался для трансляции Олимпийских игр 1964 года в Токио.

После запуска Syncom-3, 20 августа 1964 года представители 11 государств под эгидой США подписали соглашение о создании международного консорциума спутниковой связи «Интелсат». 6 апреля 1965 года консорциум запустил свой первый спутник Intelsat-1 (его символично назвали «ранней пташкой» — Early Bird). Это был первый коммерческий телекоммуникационный спутник. Оснащенный двумя транспондерами (транспондер — комплект приемопередающего оборудования на один канал) для работы в канале шириной 25 МГц, он поддерживал одновременно один телевизионный канал и 480 голосовых каналов. Сегодня в консорциум «Интелсат» входят 134 государства, он обеспечивает 2/3 всех международных телефонных разговоров, не считая других телекоммуникационных приложений. Современный спутник Intelsat VI обладает пропускной способностью 80 тыс. голосовых каналов.

23 апреля 1965 года с третьей попытки был выведен на орбиту и начал успешно работать первый отечественный спутник связи «Молния-1», положив тем самым начало региональным ССС и спутниковым группировкам на высокоэллиптических орбитах.

Так зарождались современные ССС. С тех пор в течение 40 лет процесс этот стремительно развивается. О каждом отдельном направлении ССС можно писать книгу. Поскольку такой возможности у нас нет, очень сжато расскажем, что же представляют собой спутниковые системы связи.

2.4.2 Виды орбитальных группировок.

Геостационарные орбиты

Многие важнейшие свойства ССС напрямую зависят от высоты и типа орбит КА, а также от их числа в составе орбитальной группировки. Высоты орбит не могут быть произвольными. Первое ограничение связано с тем, что наша планета окружена двумя облаками ионизированных частиц, так называемыми поясами Ван-Аллена. Эти частицы удерживаются благодаря магнитному полю Земли. Пояса Ван-Аллена расположены в экваториальной плоскости. Первый пояс — на высотах от 1,5 до примерно 4 тыс. км, его «толщина» около 600 км. Второй пояс Ван-Аллена расположен на высотах 13–19 тыс. км и занимает области порядка 500 км по обе стороны от плоскости экватора, «толщина» — 1000 км. Уровень ионизационного излучения в области поясов Ван-Аллена — порядка 10 тыс. имп./с, что по крайней мере на порядок выше прилегающих зон. Очевидно, что КА в этих районах лучше не летать. Поэтому выделяют четыре основные градации высот орбит: низкоорбитальные группировки (LEO—low earth orbit) с высотами порядка 500–2000 км, средневысотные группировки (MEO — medium earth orbit) с высотами 5–15 тыс. км (между первым и вторым поясами Ван-Аллена), геостационарные орбитальные группировки (GEO) на фиксированной высоте 36 тыс. км и орбитальные группировки на высокоэллиптических орбитах (HEO). В последнем случае орбита представляет сильно вытянутый эллипс с Землей в одном из его фокусов. Высоты в апогее (наивысшей точке), как правило, составляют порядка 7–70 тыс. км, в перигее — 400–500 км. Каждая из этих орбитальных позиций обладают своими достоинствами и недостатками.

О важности GEO еще в 1945 году писал Артур Кларк. Как и следует из названия, период обращения спутника на этой орбите строго соответствует периоду вращения Земли. Нетрудно заметить, что возможна всего одна такая орбита, расположенная в экваториальной плоскости Земли. Простейший расчет показывает, что ее радиус 42 164 км (или высота 35 768 км от поверхности Земли), соответственно, протяженность 264 790 км. КА на геостационарной орбите неподвижно висит над заданной точкой Земли, поэтому GEO еще называют геосинхронной орбитой. Очевидно, что трех спутников на GEO достаточно, чтобы, по крайней мере, один из них был в зоне прямой видимости в любой точке Земли. Неудивительно, что с 1964 года нашлось немало желающих разместить свои спутники на GEO. Проблема в том, что близко расположенные спутники, работающие в одинаковых частотных диапазонах, будут серьезно мешать работе друг друга. Ситуацию нужно было урегулировать, для чего и был организован Международный комитет по регистрации частот (IFRB — International Frequency Registration Board). Эта организация занимается вопросами распределения позиций (точек стояния) на GEO. Стандартный шаг точек стояния — 1° . Однако допускается интервал и $0,5^\circ$, если частотный диапазон соседних спутников существенно различен. Таким образом, всего возможно не более 720 точек стояния. В одной точке стояния разрешено располагать несколько спутников, если они принадлежат одной стране или компании.

Достоинства КА на GEO связаны с их неподвижностью относительно Земли и хорошей видимостью. В результате отпадает необходимость перестройки

антенн наземных станций, нет изменения частот в связи с эффектом Доплера и т. д. Но у этой медали есть и обратная сторона. Прежде всего, из-за столь высокой орбиты возникает более чем ощутимая задержка распространения сигнала на трассе «Земля — спутник — Земля» — никак не менее 500 мс $36000 \text{ (км)} \times \times 2/300000 \text{ (м/с)} = 0,48 \text{ с}$, реально — от 600 мс . Напомним, в телефонии задержка распространения сигнала не должна превышать 250 мс , иначе это становится заметным. С задержками $500\text{--}600 \text{ мс}$ не справятся никакие системы компенсации экосигнала. Высокая орбита существенно повышает требования к мощности передатчиков и чувствительности приемников, к размерам приемо-передающих антенн. Для того чтобы спутник постоянно находился в определенной точке геостационарной орбиты, требуется периодическая коррекция его полета. Наконец, чтобы вывести спутник на высокую орбиту, нужен достаточно мощный носитель. Достаточно сказать, что долгое время в СССР не могли вывести спутник на геостационарную орбиту именно из-за проблем с мощными ракетами-носителями. В США дело обстоит лучше не только из-за более чем десятилетнего технологического опережения в создании мощных носителей, но и из-за лучшего географического положения (ближе к экватору) стартовой позиции. Использование же мощного носителя неизбежно приводит к удорожанию запуска и всей ССС в целом, тем более что срок жизни КА ограничен (у современных КА — $10\text{--}15 \text{ лет}$).

Еще один недостаток КА на GEO — в полярных районах (в широтах выше $76,5^\circ$) они становятся практически невидимыми. А по крайней мере для России наличие связи в этих областях достаточно актуально. Альтернативой в данном случае может стать так называемая высокоэллиптическая орбита.

2.4.3 Эллиптические, средневысотные и низкие орбиты

Эллиптические орбиты характеризуются углом наклона плоскости орбиты к плоскости экватора, координатами узлов орбиты (точек ее пересечения с плоскостью экватора) и эксцентриситетом (отношением большой и малой осей эллипса). Видимо, первым спутником связи на эллиптической орбите стал американский Relay-1, запущенный 13 декабря 1962 года (хронологически второй активный ретранслятор на орбите). Апогей его орбиты достигал $7,4 \text{ тыс. км}$, перигей — $1,3 \text{ тыс. км}$.

Период обращения КА на эллиптических орбитах выбирается кратным суткам. Так, у отечественной ССС «Молния» высота орбиты в апогее достигает 40 тыс. км , в перигее — 460 км , наклонение — $63,5^\circ$, период обращения КА — 12 часов . Скорость движения КА в апогее значительно ниже, чем в перигее, поэтому время его видимости в требуемом регионе существенно выше, чем у спутника на круговой орбите. КА ССС «Молния» виден с территории России $8\text{--}10 \text{ часов}$, т. е. всего трех спутников достаточно для обеспечения непрерывной связи. Отметим, что из-за неоднородности гравитационного поля Земли наклонение эллиптической орбиты может быть только $63,4^\circ$ (и симметрично ей $116,6^\circ$). Иначе обеспечить стабильное положение орбиты КА проблематично. Также следует учитывать, что КА на НЕО неизбежно пересекают пояс Ван-Аллена. Еще один недостаток, свойственный всем негеостационарным ССС, — доплеровское смещение частоты из-за различной с точки зрения земного наблюдателя скоро-

сти движения спутника, а в случае НЕО у КА существенно различается еще и собственная скорость.

Средневысокие орбиты располагаются между первым и вторым поясами Ван-Аллена на высотах 5–15 тыс. км. Зона охвата таких КА существенно ниже, чем у КА на геостационарных или высокоэллиптических орбитах. Кроме того, время пребывания в зоне радиовидимости наземной станции у КА на МЕО составляет 1,5–2 часа. Поэтому для организации глобальных ССС необходимо порядка 10 КА. Из наиболее известных ССС (проекты) на МЕО можно назвать региональную систему ICO (10 активных КА + 2 резервных, высота орбиты 10 355 км, наклонение 45°) и глобальные системы «Ростелсат-В» (24 КА, высота орбиты 10 360 км, наклонение 82°) и Spaceway NGSO (20 КА, высота орбиты 10 352 км, наклонение 55°).

Конечно, увеличение числа спутников в орбитальной группировке и постоянное изменение их местоположения — это существенные недостатки ССС на МЕО по сравнению с геостационарными группировками. Однако они же превращаются в достоинства. Прежде всего существенно сокращаются задержки распространения сигнала (максимум 130 мс), что позволяет использовать такие ССС для задач телефонии. Кроме того, если в группировке несколько спутников, по крайней мере один из них виден под углом, большим 30° , а это весьма существенно для уменьшения влияния рельефа земной поверхности. Благодаря данной особенности, а также более низким орбитам существенно снижаются требования к мощности/чувствительности приемопередающей аппаратуры.

Период обращения спутников на орбитах высотой 10 350 км составляет 6 часов, при этом в тени Земли КА пребывают лишь несколько минут. Следовательно, их солнечные батареи (основной источник энергии) практически постоянно питают бортовую сеть, что существенно упрощает систему обеспечения электроэнергией и обеспечивает срок жизни КА до 10–15 лет — практически такой же, как и у спутников на GEO (в первую очередь, за счет минимального числа циклов перезарядки аккумуляторных батарей).

Низкими считаются орбиты, лежащие ниже первого пояса Ван-Аллена, т. е. с высотами 500–2000 км. Как правило, это круговые орбиты с наклонением 0 (экваториальные), 90° (полярные) и малонаклонные орбиты. Достоинства низкоорбитальных спутниковых группировок очевидны — на один-два порядка меньшее по сравнению с МЕО и GEO расстояние до ретранслятора. В результате требования к мощности наземных передатчиков и параметрам антенн снижаются настолько, что наземные абонентские станции могут уже почти не отличаться от сотовых телефонов. Именно низкоорбитальные ССС Iridium и Globalstar стали первыми глобальными системами спутниковой персональной телефонии. Низкая высота также обеспечивает малые задержки распространения, порядка десятков миллисекунд. Однако плата за низкие орбиты — необходимость в достаточно большом числе КА в составе орбитальной группировки. Это естественно, поскольку на орбитах высотой 500–2000 км период обращения составляет 1,5–2 часа. В сочетании с малой зоной радиопокрытия это приводит к тому, что в течение одного витка КА вокруг Земли абонентская станция находится в зоне радиовидимости его транспондера несколько минут. Поэтому в состав орбитальных группировок низкоорбитальных ССС входит от 48–70 КА (288 в ССС Teledesic).

Еще один минус LEO-относительно малый срок жизни КА. Прежде всего, на столь низких орбитах КА примерно треть времени находятся в тени Земли, что требует частых перезарядок аккумуляторных батарей. На орбитах ниже 700 км плотность атмосферы еще достаточна, чтобы требовалась постоянная коррекция орбиты КА. Выше же 1500 км КА оказываются в первом поясе Ван-Аллена, что сокращает время жизни электронной аппаратуры даже при использовании специальных защитных средств. В результате срок жизни КА на LEO составляет 5–7 лет.

2.4.4 Архитектура и основные принципы работы спутниковых систем связи

Любая ССС состоит из трех сегментов: пользовательского, спутникового и сегмента управления. Пользовательский сегмент — это сеть абонентских станций. Сегмент управления включает несколько наземных станций, обеспечивающих управление и сопряжение ССС с наземными системами связи. Спутники на орбите выполняют роль ретрансляторов. В этом смысле они мало отличаются от традиционных ретрансляционных систем.

Возможны три основных направления передачи информации: Земля–КА (восходящий канал), КА–Земля (нисходящий канал) и спутник–спутник. В последнем случае бортовая аппаратура КА может выполнять функцию маршрутизации трафика. КА, как правило, поддерживает несколько независимых каналов приемопередачи (стволов или транспондеров). Современные КА могут содержать десятки транспондеров. Известно несколько схем спутниковой ретрансляции. Наиболее простая — это схема группового повторителя (прозрачная ретрансляция). Принимаемый групповой сигнал (от наземной станции) преобразуется без демодуляции в сигнал выходной частоты, усиливается и транслируется. Известны методы разделения входного канала на несколько подканалов и отдельная обработка (усиление) в каждом из них. Наконец, в современных ССС, поддерживающих пакетную коммутацию, т.е. с демодуляцией и декодированием сигнала на борту КА, применяются регенеративные схемы ретрансляции — сигнал восстанавливается до исходного потока данных, а затем вновь преобразуется в ВЧ-сигнал.

С точки зрения зоны обслуживания выделяют глобальные (охватывающие значительную часть территории Земли) и региональные ССС (т.е. оптимизированные для обслуживания определенного региона или страны). Характерный пример глобальных ССС — Intelsat, Inmarsat, региональных — «Молния», Eutelsat, Arabsat.

Согласно исторической традиции ССС принято делить на системы фиксированной спутниковой связи (СФСС) и подвижной (СПСС). Сегодня это деление достаточно условно, и речь скорее надо вести о возможной скорости передачи данных (или о широкополосной и узкополосной связи). Однако данная градация отражена в действующих нормативных документах, и игнорировать ее нельзя. Для ФСС предусмотрено несколько возможных диапазонов частот: С-диапазон (5,9–6,4 и 3,7–4,2 ГГц, восходящий и нисходящие каналы соответственно), Ku-диапазон (14,0–14,5 и 11,7–12,2 ГГц) и Ka-диапазон (27–31 и 17–21 ГГц). В последнее время начинают осваивать Q- и V-диапазоны (в районе 40 и 50 ГГц). Для

межспутниковых каналов используют диапазоны 54–58 и 59–64 ГГц. Наиболее развитая система ФСС несомненно Intelsat. Основанная в 1964 году как международная организация, Intelsat в 2001 году стал частной компанией. К началу 2004 года в ее орбитальную группировку входило 28 действующих КА (и два резервных) на геостационарной орбите. ССС Intelsat связывает пользователей более чем в 200 странах мира.

Формально СПСС должны работать в L- и S-диапазонах частот (1,53–2,7 МГц и 2,7–3,5 ГГц) и ниже 1 ГГц. Но развитие микроэлектронных технологий позволяет сегодня делать вполне мобильными и абонентские станции в Ka-диапазоне. Крупнейший оператор СПСС — это основанный в 1979 году международный консорциум Inmarsat, в 1999 году преобразованный в частную компанию. Он располагает орбитальной группировкой на GEO в составе 5 КА Inmarsat I-3 третьего поколения (один — резервный), образующей глобальную систему мобильной связи. До этого в группировку входили 4 КА Inmarsat I-2 второго поколения (запуски в 1990–1992 годах). ССС Inmarsat работает в L- и S-диапазонах и поддерживает несколько стандартов предоставления услуг как телефонной связи, так и низкоскоростной передачи данных. Основной недостаток системы Inmarsat — удаленность ее КА, находящихся на GEO, и связанный с этим комплекс проблем. Даже специально разработанный терминал стандарта Inmarsat Mini-M весит около 3 кг и по размерам соответствует ноутбуку.

В начале–середине 90-х годов появилось несколько десятков проектов создания низко- и среднеорбитальных ССС, как для задач телефонии, так и для широкополосной передачи данных (к 1977 году в ИТУ поступило около 50 таких проектов). В области телефонии до практической реализации дело дошло лишь у двух глобальных низкоорбитальных систем — Iridium и Globalstar (персональные спутниковые телефоны). Первая пережила банкротство и медленно возвращается к жизни, вторая пребывает в состоянии банкротства. Та же судьба постигла и ССС на МЕО — ICO. Наиболее амбициозным низкоорбитальным проектом спутниковой СПИ была система Teledesic. Предполагалось, что эта система с орбитальной группировкой в 288 КА будет работать в Ka-диапазоне, однако подобные системы получаются очень дорогими. Наземные беспроводные СПИ развиваются столь стремительно, что конкурировать с ними низкоорбитальным ССС оказалось не под силу.

Видимо, первой услуги широкополосной передачи информации мобильным пользователям станет предоставлять Inmarsat, в ближайших планах которого — запуск на GEO спутников четвертого поколения Inmarsat I-4, обеспечивающих обмен с абонентскими терминалами на скоростях до 432 кбит/с и совместимость с системами сотовой связи третьего поколения (первый спутник Inmarsat I-4 успешно запущен 11 марта 2005 года).

Упомянем еще об одном виде низкоорбитальных ССС, так называемых little-LEO. Это ССС, работающие в частотном диапазоне ниже 1 ГГц и выполняющие функцию передачи небольших объемов информации вне реального времени. Характерный пример — системы «Гонец» и Orbcomm. КА систем «Гонец» — это ретрансляторы с зоной обслуживания диаметром примерно 5 тыс. км (высота орбиты 1500 км, наклонение 82,5°). Если принимающий абонент находится вне зоны обслуживания, информация запоминается и передается при пролете КА

через его регион. Основное назначения таких систем — электронная почта и пейджинг. Они относительно просты, что существенно снижает стоимость оборудования и услуг.

В последнее десятилетие чрезвычайно мощное развитие получили сети, использующие геостационарные КА и наземные станции с очень малой апертурой антенн (VSAT — Very Small Aperture Terminal) и их разновидность USAT — станции с ультрамалой апертурой антенн (диаметром менее 0,5 м). VSAT-системы широко используются при создании корпоративных CCC. Такие системы за счет относительной простоты и невысокой стоимости наземных станций (диаметр антенн порядка 1,5–1 м в Ku-диапазоне и 0,5 м в Ka-диапазоне) позволяют достаточно просто объединить удаленные друг от друга локальные объекты в единую сеть передачи информации со скоростями обмена порядка 2 Мбит/с и выше. И если бы не особенности национального регулирования связи, ряд из них вполне можно было бы отнести к подвижным системам.

Архитектура VSAT-сетей может быть произвольной: точка-точка, каждый с каждым, звезда. Однако для построения распределенных сетей передачи данных наиболее приемлема последняя технология. Она подразумевает наличие одной центральной наземной станции с большой антенной и ряд удаленных терминалов. Связь организуется через центральную станцию в режиме разделения времени. Используют и технологию «каждый с каждым», при этом каналы, как правило, не фиксированы и предоставляются по запросу (механизм DAMA). С точки зрения спутникового сегмента отличие технологии VSAT — зона обслуживания не покрывается одним глобальным лучом, а разбита на ряд сегментов, охватываемых рядом относительно узких лучей. Такой метод поддерживают практически все современные КА на GEO. Отметим, что на фоне неудач низкоорбитальных спутниковых систем передачи данных рынок VSAT-систем успешно развивается, в том числе и в нашей стране. В качестве характерной VSAT-системы назовем принадлежащую ОАО «Газпром» систему «Ямал».

2.4.5 Методы множественного доступа в CCC

Один из основных вопросов построения телекоммуникационных сетей — доступ множества абонентских (терминальных) устройств к общему коммуникационному ресурсу, в случае CCC — к транспондерам. Существует несколько методов мультиплексирования каналов.

Исторически первый метод мультиплексирования — это частотное разделение каналов (FDM), в том числе дуплексных (FD). Самая простая его разновидность — фиксированное назначение диапазона каждой станции (FAMA — fixed-assignment, multiple access). Каждому каналу присваивается свой фиксированный диапазон (своя несущая). В результате для того, чтобы установить дуплексное соединение между двумя станциями, необходимо задействовать четыре частотных канала (т. е. два транспондера). Данный механизм существенно упрощает бортовую аппаратуру КА, ретранслятор работает в режиме повторителя. Однако его ресурс расходуется достаточно неэффективно, фактически число возможных соединений точка-точка между наземными станциями равно числу транспондеров у КА. Типичный пример — CCC, работающие в C-диапазоне (по-

давлиющее большинство стволов отечественных геостационарных ССС). Ширина полосы восходящего/нисходящего канала при этом 500 МГц. Как правило, КА поддерживает 12 транспондеров, ширина канала каждого — 36 МГц (плюс защитный интервал 4 МГц между каналами и по 10 МГц на краях диапазона 500 МГц). Число каналов можно удвоить, используя ортогональную поляризацию сигнала: 12 нечетных транспондеров работают с вертикальной поляризацией антенн, 12 четных — с горизонтальной. При этом центральная несущая вертикально поляризованных каналов смещена на 20 МГц относительно горизонтально поляризованных. Метод FDM/FD/FAMA хотя и наиболее простой в реализации, но и наиболее неэффективный, особенно когда увеличивается число наземных терминалов.

Дальнейшим развитием множественного доступа с частотным разделением каналов (FDMA) стал метод SCPC (single channel per carrier) — один канал на несущую. Речь идет в данном случае о телефонном канале шириной 4 кГц. Суть его в том, что весь частотный диапазон делится, например, на субканалы 4 кГц, каждый со своей независимо модулируемой несущей. Распределение подканалов между станциями происходит на основе метода DAMA (demand-assignment, multiple access) — множественный доступ с назначением канала по запросу. Это означает, что перед началом трансляции каждая станция сообщает центральной (управляющей), что она хочет установить соединение с другой станцией. Для служебных сообщений выделяется специальный сигнальный канал, с помощью которого происходит распределение подканалов между станциями, так, чтобы не было двух станций, одновременно работающих в одном частотном подканале. Скорость передачи данных в каждом подканале зависит от его ширины, механизма кодирования и вида модуляции. Впервые механизм FDMA/DAMA был реализован компанией Comsat в спутниках Intelsat серий IV A и V. В этой системе в сигнальном канале обмен происходит в режиме доступа с разделением времени (TDM). Информация передается в циклически повторяющихся кадрах длительностью 50 мс. Кадр разбит на 50 тайм-слотов по 1 мс. Каждый тайм-слот закреплен за определенной наземной станцией (в системе Intelsat их не более 50). В течение тайм-слота наземная станция может передать на частоте служебного канала 128 бит — запросы на предоставление голосового канала или сообщение об освобождении канала.

Исторически следующим после FDMA стали использовать механизм множественного доступа с разделением каналов по времени (TDMA — time-division multiple access). Принцип его прост: вся передача происходит в циклически повторяющихся кадрах (фреймах), разделенных на интервалы (тайм-слоты) равной длительности. Каждому передатчику назначается (на постоянной или временной основе) определенный тайм-слот. Основная проблема при этом — не потерять кадровую синхронизацию, для чего в начале каждого кадра передается определенная синхропоследовательность. В свое время Европейская конференция почтовых и телекоммуникационных ведомств СЕРТ приняла как стандарт структуру базового TDMA-кадра для передачи голосовой информации. Предполагается, что после оцифровки и кодирования поток голосовых данных в каждом канале представляет последовательность 8-разрядных выборок с частотой 16 кГц. Один TDMA-кадр обслуживает 16 независимых голосовых каналов. Выборки каждого

их них (по 8 бит) последовательно формируют подкадр из 128 бит. Скорость передачи (с учетом частоты выборок 16 кГц) 2,048 Мбит/с. 32 последовательных подкадра образуют кадр длительностью 2 мс. Это означает, что каждая наземная станция может передавать данные в одном голосовом канале порциями по 32 8-разрядные выборки каждые 2 мс. Отметим, что реальное время передачи базового TDMA кадра в 59 раз меньше (33,9 мкс) — скорость передачи наземной станцией базового TDMA-кадра СЕРТ составляет 120,832 Мбит/с.

Достоинство технологии TDMA перед FDMA прежде всего в том, что в транспондере все время присутствует одна модулированная несущая, т. е. существенно снижаются межканальные интермодуляционные помехи. Упрощается и аналоговая часть аппаратуры (нет необходимости отдельно обрабатывать множество частотных каналов). Технология TDMA гораздо лучше подходит для множественного доступа при передаче цифровых данных.

Более универсальным и перспективным методом множественного доступа является технология кодового разделения каналов (CDMA). Основные ее достоинства: невысокая пиковая мощность сигнала, гибкость перестройки каналов и выбора полосы канала, простота перехода с канала на канал, возможность работы нескольких станций в одном частотном диапазоне. Метод этот относительно недавно нашел применение в СССР (впервые — в системе Omnitrac). Характерные примеры использующих его систем — СССР Globalstar и Ellipso. Отметим, однако, что именно на CDMA (в различных вариациях этой технологии) основываются многие проекты перспективных СССР.

Еще один перспективный метод множественного доступа — мультиплексирование посредством ортогональных несущих (OFDM). Он требует относительно сложных средств цифровой обработки сигнала, поэтому до определенного момента и не получил должного распространения. Однако именно на основе данной технологии построен стандарт цифрового телевизионного вещания DVB, в том числе и его спутниковый вариант DVB-S.

В заключение беглого обзора СССР отметим, что они все еще остаются достаточно дорогим удовольствием. Именно поэтому ряд технически восхитительных проектов так и не был реализован (например, Teledesic — глобальная космическая сеть передачи данных). По-видимому, их время пока не пришло.

2.5 Оптические атмосферные линии связи и сети

Как отмечалось в предыдущих разделах, в вычислительных системах и сетях активно используются широкополосные радиосредства, обеспечивающие высокоскоростную связь и быстрое развертывание между узлами сетей. Основным недостатком такого канала является необходимость получения разрешения на используемые диапазоны частот, ограничения по диапазону частот и мощности передатчика.

Оптические атмосферные линии связи свободны от этих недостатков и придают вычислительным, телекоммуникационным системам и сетям такие качества, как:

- невосприимчивость к электромагнитным помехам;
- высокая скорость передачи;

- низкая удельная себестоимость бита передаваемой информации;
- совместимость по техническим характеристикам оптоэлектронных приборов и устройств с микроэлектронными схемами и устройствами.

Применение их в вычислительных сетях, безусловно, перспективно. Поэтому особый интерес представляет использование устройства, состоящего из оптического передатчика — оптической среды (атмосферы), — фотоприемника, то есть оптоэлектронной атмосферной линии передачи информации (при наличии каналаобразующих устройств она выполняет функцию канала). Оптический передатчик обычно включает инфракрасный излучающий диод или полупроводниковый лазерный диод и схему управления, фотоприемник — pin-фотодиод и усилитель выходного сигнала. Так как такой канал вызывает интерес у потребителей различных областей техники, то его часто называют инфракрасной линией связи или лазерной линией связи и т. д. (хотя в них всегда происходят присущие оптоэлектронным приборам, схемам и системам физические преобразования «электричество свет-электричество» или их комбинации).

Многие из недостатков электромагнитных приборов, устройств и систем (особенно возможность подавления абонента мощным электромагнитным излучением, грозовыми разрядами, ядерным излучением и т. д.) в таком канале устранены.

Появление полупроводниковых лазерных диодов на основе арсенида галлия, легированного алюминием, позволило получить оптоэлектронные атмосферные каналы с длиной волны излучения 850–870 нм. Такие диоды позволяют строить оптоэлектронные атмосферные каналы со следующими характеристиками:

- высокая скорость передачи информации, до 155 Мбит/с;
- вероятность ошибки при передаче информации 10^{-9} ;
- мобильность связи;
- повышенная секретность связи.

Но использование лазерных диодов накладывает следующие ограничения на оптоэлектронный атмосферный канал (ОАК):

- чтобы обеспечить экологическую безопасность лазерных атмосферных линий (ЛАЛ), в оптических передатчиках таких линий используют полупроводниковые лазерные диоды с мощностью до 50 мВт;
- с другой стороны, чтобы добиться приемлемой дистанции передачи оптического сигнала от передатчика до приемника ЛАЛ — 1–2 км (т. е. обеспечить приемлемую интенсивность излучения в телесном угле), требуется сформировать узконаправленный луч с углом расхождения в несколько единиц или даже долей мрад;
- это, в свою очередь, предъявляет жесткие требования к стабилизации передатчика (т. е. заставляет разработчика или делать массивное основание, или ставить передатчик на гиросплатформу). Но, сужая луч и стабилизируя передатчик ЛАЛ, не всегда удается обеспечить его стабильную работу, так как коэффициент преломления реальной атмосферы непостоянен. Поэтому необходимо увеличивать расходимость луча ЛАЛ и, соответственно, снижать дистанцию передачи;
- все это, вместе взятое, заставляет разработчика использовать сложные системы стабилизации ЛАЛ, что, в свою очередь, повышает его стоимость.

Рассмотрим далее особенности конструирования и применения инфракрасных атмосферных каналов в компьютерных сетях и перспективы их дальнейшего развития.

Одна из простейших топологий симплексного инфракрасного атмосферного канала (ИАК) на основе инфракрасного диода и ррп-фотодиода представлена на рис. 2.66.

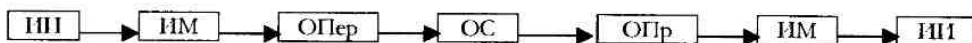


Рис. 2.66. Структурная схема симплексного инфракрасного канала передачи данных: ИИ — источник информации; ИМ — интерфейсный модуль; ОПер — оптический передатчик (схема управления током, протекающим через инфракрасный излучающий диод, и сам инфракрасный диод, линзовая система формирования выходного оптического луча); ОС — оптическая среда (атмосфера); ОПр — оптический приемник (ррп-фотодиод, схема усилителя-формирователя выходного сигнала и линзовая система приема оптического сигнала)

ИАК представляет собой устройства, преобразующие электрическую форму представления информации в оптическую и обратно. Оптическая среда ИАК (атмосфера), рассогласование оптических параметров передатчика и приемника вносят значительное ослабление сигнала. Общее затухание информационного оптического сигнала ИАК складывается из двух основных составляющих:

- 1) затухание оптического сигнала из-за рассогласования выходных оптических параметров передатчика и входных параметров приемника;
- 2) затухание за счет поглощения квантов света на парах воды, молекулах углекислого газа, аэрозольных частицах.

В схематичном виде ИАК можно представить (см. рис. 2.67) в виде расположенных на одной оптической оси оптического передатчика (состоящего из излучающего инфракрасного диода и линзы, формирующей узконаправленный луч) и оптического приемника (состоящего из оптической линзы, в фокусе которой расположена непосредственно площадка фотодиода).

На рисунке и далее введены следующие условные обозначения:

ИИ — источник излучения, в нашем случае инфракрасный излучающий диод с длиной волны излучения $\lambda = 0,85 \div 0,87$ мкм;

$L_{\text{пер}}$ — линза оптического передатчика;

$L_{\text{пр}}$ — линза оптического приемника;

Оср — оптическая среда (атмосфера);

$ОС_{\text{пер}}$ — оптическая система передатчика (ИИ + $L_{\text{пер}}$);

$ОС_{\text{пр}}$ — оптическая система фотоприемника ($L_{\text{пр}}$ + ФПУ);

ФПУ — фотоприемное устройство (Si-фотодиод + усилитель-формирователь);

$D_{\text{л.пер}}$ — диаметр линзы передатчика;

$D_{\text{л.пр}}$ — диаметр линзы приемника;

$D_{\text{итп}}$ — диаметр оптического «пятна» от оптического луча передатчика в плоскости линзы фотоприемника;

$F_{\text{пер}}$ — фокусное расстояние линзы передатчика;

$F_{\text{пр}}$ — фокусное расстояние линзы приемника;
 A — размер источника излучения;
 $Q_{\text{расх}}$ — угловая расходимость на выходе линзы передатчика; $Q_{\text{расх}} = A/F_{\text{пер}}$
 $\alpha_{\text{тах}}$ — максимальный угол выхода оптического излучения из кристалла излучающего диода;
 $\alpha_{\text{охв}}$ — угол охвата оптического излучения излучающего диода линзой передатчика;
 L — длина трассы.

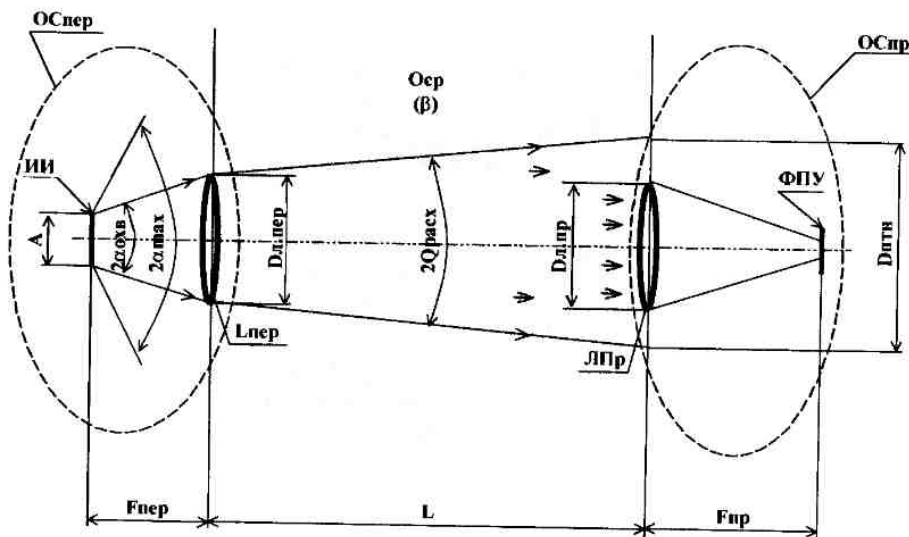


Рис. 2.67. Оптическая схема симплексного инфракрасного канала передачи данных

Расчет оптической системы сводится к определению (см. рис. 2.67) доли излученной ИК мощности, попадающей на ФПУ. Очевидно, что чем эта доля больше, тем эффективнее передача оптического сигнала от передатчика к приемнику. Падение информационного оптического сигнала за счет расхождения оптического пучка упрощенно определяется в соответствии с законами геометрической оптики отношением площади оптического пятна в плоскости линзы фотоприемника к площади линзы фотоприемника, т.е. на площадь линзы фотоприемника падает часть инфракрасного излучения, пропорциональная отношению

$$A_{\text{пр}} = (D_{\text{птн}}/D_{\text{л.пр}})^2.$$

В соответствии с законами геометрической оптики:

$$D_{\text{птн}} = L \cdot Q_{\text{расх}}.$$

Конструктивно линия связи инфракрасного диапазона представляет собой два одинаковых блока, находящихся в климатических защитных кожухах для установки вне помещений. Каждый блок имеет разъемы для непосредственного подключения к кабельной сети здания через стандартные интерфейсы, к которым относятся: V.35, G.703, Ethernet (Half и Full Duplex), ATM-155. Кроме того, существуют модели с интерфейсом в виде оптоволоконного разъема, которые

поддерживают такие протоколы, как Fast Ethernet, E1, E3, FDDI, ATM. В каждом блоке находится приемник и передатчик. Передатчик представляет собой излучатель на основе импульсного полупроводникового лазерного диода (иногда обычного светодиода). Приемник в большинстве случаев имеет в своей основе скоростной pin-фотодиод или лавинный фотодиод.

И приемник, и передатчик снабжены мощными объективами, благодаря чему луч обладает малым углом расходимости. Передаваемая информация кодируется короткими импульсами излучения. Лазерные приемопередатчики осуществляют двустороннюю связь, то есть два параллельных луча распространяются в противоположных направлениях от передатчиков к приемникам. Обязательным условием применения ЛАЛ является наличие прямой геометрической видимости между абонентами, для обеспечения которой приемопередающие блоки устанавливаются вне помещений, например на крышах зданий, балконах последних этажей и т. п. Лазерные линии обычно рассчитаны на функционирование при температуре окружающей среды от -60 до $+50^{\circ}\text{C}$.

Устанавливаются блоки ЛАЛ таким образом, чтобы оптические оси приемо-передатчиков совпадали. Сложность монтажных работ по установке и настройке ЛАЛ зависит от конкретной модели устанавливаемой линии. В целом время, затрачиваемое на установку и настройку ЛАЛ, варьируется от 30 минут до нескольких часов.

Важнейшим свойством ЛАЛ является высокая степень защиты канала от несанкционированного доступа. Защищенность канала является следствием самой природы ЛАЛ, а не обеспечивается какими-либо специальными методами. Осуществить перехват канала технически весьма трудно в силу острой направленности луча и уникального для каждой модели метода кодирования информации импульсами излучения.

Одним из основных факторов, определяющих возможность применения ЛАЛ, является устойчивость связи при воздействии помех различного происхождения, как естественного, так и искусственного, в частности снега, тумана, дыма и других явлений, снижающих прозрачность атмосферы.

Характеристикой воздействия атмосферы на связь можно считать средний процент нерабочего времени, то есть период, в течение которого связь отсутствовала. Анализ типовых условий применения ЛАЛ в городах показал, что дальность связи в основном находится в диапазоне от 1 до 3 км. С увеличением расстояния возрастают помехи, и на дальности 10 км неблагоприятные для связи условия наблюдаются в течение срока, длительность которого составляет 1,5–2% от общего времени работы. Причем снегопад является виновником отсутствия связи в половине всех случаев, на долю тумана приходится 30% и дыма — 20%.

Основными параметрами, которые нужно учесть при организации связи, являются дальность, используемый интерфейс (то есть скорость связи, протокол, тип кабельного соединения и пр.) и, главное, цена. ЛАЛ зарубежных производителей находятся в той же ценовой нише, что и радиорелейные линии, то есть в зависимости от технических характеристик их стоимость в большинстве случаев лежит в пределах от 10 до 20 тыс. долл. Что касается цен на ЛАЛ отечественного производства, то они заметно ниже. Следует отметить, что стоимость оборудования ЛАЛ значительно возрастает при увеличении дальности связи. В настоя-

щее время производится целый ряд моделей ЛАЛ. К наиболее известным зарубежным компаниям, выпускающим лазерные линии связи, относятся: Canon Inc, Laser Communications Inc, Jolt Communications Ltd, Freebird Communications Ltd, Modular Technology PLC, A.T. Schindler Communications. Среди отечественных разработчиков следует отметить Воронежский НИИ связи, НИИ ПП, Государственное предприятие «Полюс», Государственный Рязанский приборный завод, Институт проблем передачи информации РАН.

Наиболее интересной ЛАЛ в плане практической реализации уникальных технологий и, соответственно, наиболее дорогой является Canobeam II — продукт известной во многих других областях фирмы Canon. Canobeam II поддерживает связь со скоростью 155 Мбит/с (АТМ или 4 канала видео) при дальности до 4 км. Угол расходимости луча передатчика сделан очень узким — порядка 4 угловых минут, за счет чего средняя мощность излучения не превышает 10 мВт при дальности связи до 4 км. При таком узком угле луча связь не может быть устойчивой из-за ветра и суточных смещений зданий. Эта проблема решена путем отслеживания направления на следующий приемопередающий блок и автоматической корректировки оптической оси излучения. Корректировка осуществляется путем смещения специальных зеркал внутри блока. Другой важной особенностью Canobeam является специальное расширение лазерного луча непосредственно в передатчике, вследствие чего выходящий луч имеет пониженную мощность на единицу площади в непосредственной близости от передатчика и удовлетворяет стандарту безопасности IEC 825. Разумеется, все уникальные черты этой ЛАЛ отразились на ее цене, которая составляет 110 тыс. долл.

Большинство зарубежных моделей ЛАЛ имеют диапазон рабочих температур от -30 до $+50^{\circ}\text{C}$, что делает их неприемлемыми для многих районов нашей страны. Распространенной задачей беспроводного соединения с помощью ЛАЛ в России является связь между полевыми точками в районах Севера, где вечная мерзлота препятствует прокладке кабеля, северные сияния создают помехи для радиосвязи, а температура зимой нередко опускается до -60°C . Некоторые производители, например Laser Communications, предлагают помещать приемопередающие блоки в специальный саркофаг. Разумеется, это вызывает дополнительные финансовые затраты. Приемлемый перечень поддерживаемых интерфейсов, конкурентоспособные цены и способность работать в более жестких климатических условиях позволяют успешно применять отечественные лазерные линии. Дальность связи большинства отечественных моделей ЛАЛ лежит в пределах от 200 м до 10 км. Они способны передавать данные с максимальной скоростью 34,368 Мбит/с. В числе недавно запущенных в производство новинок можно отметить линию связи на 155 Мбит/с, разработанную Государственным Рязанским приборным заводом. Дальность этой ЛАЛ составляет 3 км. Диапазон рабочих температур — от -60 до $+50^{\circ}\text{C}$ — не обеспечивается ни одной зарубежной ЛАЛ. Для монтажа и настройки эти линии снабжены юстировочной платформой и оптическим прицелом.

ГЛАВА 3

Математические основы передачи информации

В последующих трех главах будут описаны основы области знаний, которая имеет несколько названий, иногда обозначающих одно и то же, иногда смежные или пересекающиеся области. Это теория информации, теория передачи информации, теория связи, теория сигналов и т.д. Название не столь важно. Важно то, что фундаментальные математические основы передачи информации интенсивно развивались в течение последних 50 лет прошлого столетия, начиная с работ Шеннона и Котельникова. С задержкой в несколько десятков лет результаты этих исследований стали реализовываться на практике, и сегодня мы присутствуем на том временном этапе, когда любые, даже самые фантастические элементы теории передачи информации реализуются в реальных системах. Как правило, реализация этих идей происходит на физическом уровне (РЧУ) модели взаимодействия открытых систем (ОСИ) и иногда на подуровне управления доступом к каналу (МАС) канального уровня. Именно поэтому сначала рассматриваются сами фундаментальные основы теории передачи информации, а затем международные стандарты, на основе которых строятся реальные системы, и сами реальные системы.

3.1 Структура и функциональная схема систем передачи дискретной информации.

Описание элементов систем (кодеки, модемы, каналы)

Блок-схема любой системы передачи дискретной информации «точка-точка» показана на рис. 3.1. Различным видам кодеров на передающей стороне соответствуют «обратные» декодеры. Кодирование используется для сжатия информации, криптографической защиты, а также повышения помехоустойчивости за счет внесения избыточности на передающей стороне. Выход корректирующего кодера и вход соответствующего декодера образуют вход и выход дискретного канала связи.

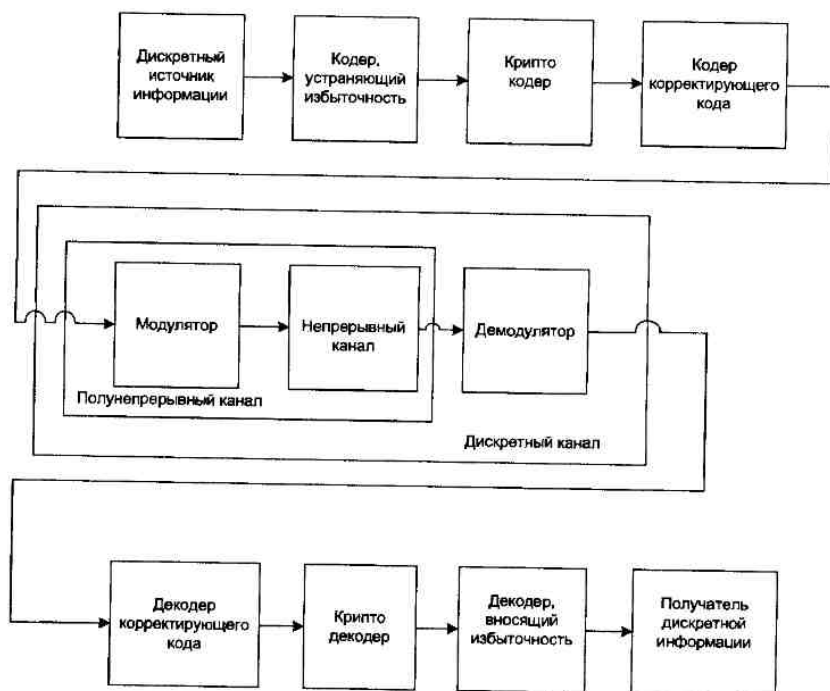


Рис. 3.1. Блок-схема системы передачи дискретной информации «точка-точка»

В общем случае дискретный канал связи является математическим объектом с I входами и J выходами с определенными переходными вероятностями. Если детализировать физическое содержание дискретного канала, то окажется, что он состоит из последовательной цепочки «модулятор—непрерывный канал—демодулятор», причем цепочка «модулятор—непрерывный канал» образует полунепрерывный канал. Непрерывный канал является каналом с непрерывным входом и непрерывным выходом, а полунепрерывный канал — с дискретным входом и непрерывным выходом. Позже будет видно, что непрерывный канал используется для синтеза сигнально-кодовых конструкций, а полунепрерывный канал — для декодирования с мягким решением. Важно также понимать, что все рассмотренные каналы являются математическими моделями соответствующих физических каналов связи. Реальная система передачи дискретной информации содержит большое число узлов и функциональных элементов, не изображенных на рис. 3.1, тем не менее без которых система не функциональна. В первую очередь, это устройства синхронизации по несущей, тактовой и блоковой частоте, без которых демодуляция и декодирование сигналов невозможны. Кроме того, это задающие генераторы, системы управления, передатчики, приемники, антенны в случае радиосистем и многие другие элементы. Однако данное рассмотрение мы ограничим узлами, изображенными на рис. 3.1. Если рассматриваемая система является непрерывной, то при помощи теоремы Котельникова, рассмотренной

ниже, система из непрерывной превращается в дискретную. Цепочка этих преобразований показана на рис. 3.2.

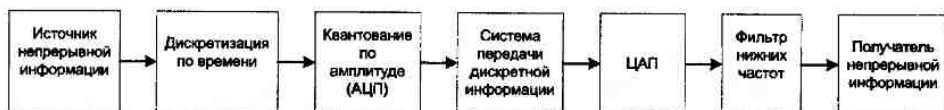


Рис. 3.2. Схема преобразований непрерывной системы связи в дискретную

Таким образом, дискретные системы передачи покрывают все системы, которые будут рассмотрены ниже.

3.2 Модели и типы каналов широкополосного доступа

При рассмотрении необходимо различать реальные физические каналы связи, возникающие при широкополосном доступе, и соответствующие им модели. В зависимости от среды передачи широкополосный доступ может быть проводным и беспроводным. Проводные каналы характеризуются более постоянными параметрами, чем беспроводные. Всем каналам — и проводным, и беспроводным присущи помехи и мешающие воздействия от других абонентов системы. В случае беспроводных каналов параметры последних могут существенно меняться во времени. Это может быть вызвано, во-первых, замираниями передаваемого сигнала в результате отражений во время распространения и, во-вторых, эффектом Доплера, возникающим при движении абонентов.

Все это приводит к большому разнообразию соответствующих моделей каналов, необходимых для синтеза сигналов, кодов и сигнально-кодовых конструкций. Первым приближением является рассмотрение так называемых каналов дискретного времени, когда рассматриваются только временные отсчеты сигнала в моменты времени $1, 2, \dots, k, k+1, \dots$. Эти сигналы могут возникать, например, в моменты стробирования неким стробирующим устройством. Вторым существенным ограничением является рассмотрение каналов без памяти, когда k -й отсчет не зависит от предыдущих и последующих отсчетов. Третьим существенным ограничением является ограничение мощности сигнала $P_{\text{вх}}$ на входе канала. И наконец, четвертым ограничением является аддитивный характер шума в канале. Все это приводит к каналу с отношением сигнал/шум $P_{\text{вх}}/P_{\text{ш}}$.

Если данный непрерывный канал квантуется на I состояний по входу и J состояний по выходу, получается дискретный канал без памяти с I входами и J выходами с соответствующими переходными вероятностями. На входе канала возникает одна из I букв из алфавита $A = \{a_1, a_2, \dots, a_I\}$, а на выходе одна из J букв из алфавита $B = \{b_1, b_2, \dots, b_J\}$. Переходные вероятности получить на выходе букву b_j при условии, что на входе была буква a_i , обозначаются как $p(b_j/a_i)$, $i = 1, \dots, I$, $j = 1, \dots, J$. Важно понимать, что в общем случае размеры алфавитов по входу и выходу канала не совпадают. Дискретный канал без памяти называется симметричным, если набор переходных вероятностей для одного

входа или выхода задает простыми перестановками соответствующие наборы для других входов или выходов.

Самым простым каналом называется двоичный симметричный канал, для которого алфавиты по входу и выходу совпадают и состоят из двух букв 0, 1. Переходные вероятности равны $p(0/0) = p(1/1) = 1 - p$, а $p(0/1) = p(1/0) = p$. Тогда p называется вероятностью ошибки, а $1 - p$ – вероятностью правильного приема.

Также очень прост так называемый стирающий канал, для которого имеется два входа 0, 1 и три выхода 0, 1, *. Символ * называется стиранием или отказом. Переходные вероятности равны $p(0/0) = p(1/1) = 1 - \alpha$, $p(0/1) = p(1/0) = 0$, а $p(0/*) = p(1/*) = \alpha$. Тогда α – это вероятность стирания, а $1 - \alpha$ – это вероятность правильного приема. В случае полунепрерывного канала, необходимого при мягком декодировании, выход канала обычно квантуется на несколько значений. В этом случае размер алфавита по выходу канала J существенно превышает алфавит по входу канала I . Очень большое внимание в четвертой главе будет уделяться гауссову каналу с межсимвольной интерференцией. Такой канал дискретного времени состоит из последовательного включения линейного фильтра и источника аддитивного гауссова шума. Наиболее традиционными моделями непрерывных каналов с памятью являются каналы с релейским или райсовским замиранием. Часто используют дискретные модели каналов с памятью, самыми простыми из которых являются каналы, образованные простыми цепями Маркова. В каждом из нескольких состояний канал соответствует определенному ДСК.

3.3 Система передачи дискретной информации как система широкополосного доступа к информационным ресурсам. Практические примеры систем

Ниже мы будем рассматривать специфические системы передачи дискретной информации, которые будем называть системами широкополосного доступа к информационным ресурсам. Под это определение подпадает огромное количество систем – это и проводные системы DSL (digital subscriber line) передачи дискретной информации по медным проводам, это и беспроводные системы BWA (broadband wireless access), это и системы цифрового радиовещания и телевидения, это и спутниковые системы и сети и многие другие. Системы могут быть и дуплексными (двунаправленными), и симплексными (однонаправленными), и проводными, и беспроводными. Системы могут быть предназначены для любого вида телеметрической информации, для передачи телефонии, а также трансляции видео и звука. Абоненты в системах широкополосного доступа могут быть неподвижны, подвижны и ограниченно подвижны.

Однако все системы широкополосного доступа объединяет несколько общих свойств.

- Все системы широкополосного доступа стремятся использовать среду передачи (канал связи) на пределе физических возможностей, т.е. пропускной способности, что влечет за собой применение новейших методов синтеза и приема сигналов, кодов и сигнально-кодовых конструкций.
- Все системы широкополосного доступа используют эффективные методы доступа к среде передачи и мультиплексирования сообщений в этой среде, что влечет за собой применение новейших методов множественного доступа.
- Все системы широкополосного доступа для максимизации своей эффективности используют сжатую информацию, что влечет за собой использование новейших методов кодирования источника.
- При передаче непрерывных сообщений используется их оцифровка и дискретная передача. Если на начальном этапе развития таких методов оцифровка сообщения приводила к существенному увеличению полосы сигнала, то теперь оцифрованное сообщение часто занимает полосу в три-четыре раза меньше исходного аналогового.
- И наконец, все системы широкополосного доступа, заменяя собой узкополосные и среднеспосные системы, существенно расширяют пользовательские свойства и приводят к мультисервисности услуг для абонентов.

Теперь можно рассмотреть фундаментальные основы теории передачи информации, позволяющие эффективно строить системы широкополосного доступа.

3.4 Аналоговая и цифровая передача информации

3.4.1 Информация как физическая величина

Информация — количественная мера степени упорядоченности исследуемой системы. Информацией обладают источник сигналов, приемник сигналов, канал передачи.

Сигнал — функция $s = S(t; \bar{p})$, характеризующая процесс передачи энергии от источника (передатчика) к приемнику через специально организованную среду — канал. В выражении $S(t; \bar{p})$ переменная S — энергетическая характеристика сигнала (амплитуда, фаза, частота), $\bar{p}$ — набор параметров, t — время. Очевидно, характеристика будет энергетической, если она выражает непосредственно энергию сигнала (амплитуда, квадрат амплитуды, мощность) либо функционально связана с энергией (спектральная плотность и др.). Сигнал аналоговый, если его характеристика континуальная (непрерывная) величина, и дискретный, если S принимает значения из конечного или бесконечного дискретного ряда.

Переменная S дискретна, если для любого ее значения s' , $s = s'$ существует такое значение s'' , $s'' \neq s'$, что между s' и s'' невозможно вставить никакое другое значение переменной s . Формально: пусть $s(s', s'' \dots) \in S$, тогда $(\forall s' \in S) (\exists s'' \neq$

$\neq s') (\exists s) [(s' < s < s'') \vee (s'' < s < s')]$ — кванторная предикатная запись свойства дискретности сигнала.

Связисты под информацией понимают количество разрядов (длину) $L = |\bar{a}|$ кодового слова $\bar{a} = a_{(1)}a_{(2)} \dots a_{(L)}$, $(\forall l)a_{(l)} \in \{0; 1\}$. С помощью кодовых слов кодируется определенное сообщение (сигнал) s из множества S всех возможных сообщений, $s \in S$. Только что приведенное определение информации может быть обобщено на произвольный алфавит A_N — множество (список) из N символов (состояний). Бинарный алфавит $A_2 = \{0; 1\}$ — частный случай; A_2 используют чаще других, потому что имеются электронные схемы, которые могут находиться в одном из двух устойчивых состояний. Схемы с тремя и более устойчивыми состояниями менее надежны, чем с двумя состояниями.

Энтропия H — количественная мера хаотичности исследуемой системы. Пусть H_1 — начальная энтропия приемника сигналов, H_2 — конечная энтропия того же приемника после приема сообщения s . Информация, содержащаяся в s , равна уменьшению энтропии

$$\mathbb{I}(s) = H_1 - H_2 = \Delta H \geq 0. \quad (3.1)$$

Понятие информации, принятое в физике, на наш взгляд, предпочтительнее абстрактного, используемого математиками. Математики предпочитают выводить формулу для максимальной информации $\mathbb{I}_{max} = \log N$, где N — количество равновероятных сообщений, исходя из набора заданных формальных требований (см., например, такой вывод в книге Яноши [114]). Физический подход показывает, во-первых, что информация — измеримая приборами величина, а не субъективная или абстрактная сущность, во-вторых, что она связана с энергией, т. е. с сигналом, зависит от вида и параметров последнего.

К сожалению, в научной литературе сложилась несколько странная традиция: энтропия в физике и энтропия в теории информации характеризуются по-разному (см. к примеру: Физическая энциклопедия. 1966. Т. 5. С. 533–534). Связь между двумя видами энтропий, фактически их тождество, ускользает от читателей, из-за чего возникают большие трудности понимания и оценивания информации в реальных системах связи и управления.

3.4.2 Энтропия Л. Больцмана

Энтропия H системы невзаимодействующих на расстоянии хаотически движущихся молекул определяется по удивительно простой формуле, принадлежащей Людвигу Больцману:

$$H = k_B \cdot \ln W(\mathcal{E}) \left[\frac{\text{Дж}}{K} \right], \quad (3.2)$$

где $W(\mathcal{E})$ — Вероятность $\{\text{energy} \leq \mathcal{E}\}$ — функция распределения вероятностей случайной энергии **energy**, которой обладает молекула, наугад выбранная из системы молекул, при этом для стационарной системы функция $W(\mathcal{E})$ зависит еще и от температуры T как от параметра; k_B — постоянная Больцмана (энергия,

приходящаяся на одну молекулу газа а стационарном состоянии):

$$k_B = 1,3802 \cdot 10^{-23} \left[\frac{\text{Дж}}{\text{К}} \right]$$

Уточним физический смысл функции $W(\mathcal{E})$ применительно к информационным проблемам.

Пусть имеется конечное число частиц (возможных «сообщений») N . Обозначим $n_1, n_2, \dots, n_K$ — число частиц с энергиями $\mathcal{E}_k$, $n_k = n(\mathcal{E}_k)$, $k = 1, 2, \dots, K$. Тогда

$$W(\mathcal{E}) = \frac{N!}{\prod_{k=1}^K n_k!} \quad (3.3)$$

— распределение числа размещений N частиц («сообщений») по K «ящикам» (каналам) (рис. 3.3).

Посмотрим, во что превратится выражение (3.3) при больших N , $n_k \rightarrow \infty$. По известной формуле Стирлинга $n! \sim n^n e^{-n}$ получаем для энтропии

$$H \sim k_B \left(N \ln N - \sum_k n_k \ln n_k \right). \quad (3.4)$$

Вопрос: а при каком распределении вероятностей $W(\bar{n}) = W(n(\mathcal{E}_1), n(\mathcal{E}_2), \dots, n(\mathcal{E}_K)) = W(n_1, n_2, \dots, n_K)$ энтропия (3.4) будет максимальной? Иначе, по какому закону распределены частицы («сообщения») по уровням энергии в системе, живущей по законам полного хаоса? Применительно к системе связи можно поставить вопрос по другому: какому вероятностному закону подчиняются длины сообщений абсолютно случайного источника? Для системы обслуживания (например, для приемника) этот вопрос можно поставить сходным образом: как распределены промежутки времени между прибытием последовательных сообщений в случайном канале связи или сообщений, поступающих от абсолютно случайного источника? То, что энтропия таких источников максимальна, имеет важное практическое значение: случайный источник — в известном смысле наихудший случай для приемника.

«Наихудшее» распределение, максимизирующее энтропию, оказалось экспоненциальным:

$$W(\mathcal{E}_k) = \exp(-\alpha - \beta \mathcal{E}_k), \quad (3.5)$$

при этом для n_k имеем выражение

$$n_k = N e^{-\beta \mathcal{E}_k}. \quad (3.6)$$

Как находится выражение (3.5)? Путем максимизации (3.4) при ограничениях, накладываемых на величины n_k :

$$H \Rightarrow \max$$

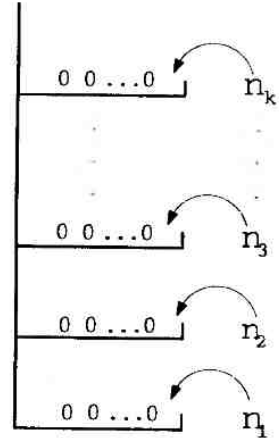


Рис. 3.3. Схема «ящиков»

$$\mathcal{E} = \sum_{k=1}^K n_k \mathcal{E}_k = \text{const} \quad (\text{закон сохранения энергии}),$$

$$N = \sum_{k=1}^K n_k = \text{const} \quad (\text{закон сохранения количества частиц}).$$

Составляется функция Лагранжа

$$F = H + \alpha \left(\sum n_k - N \right) + \beta \left(\sum n_k \mathcal{E}_k - \mathcal{E} \right)$$

и максимизируется путем дифференцирования по переменным n_k , $k = 1, 2, \dots, K$, и по переменным α и β . Приравнявая производные F к нулю и записывая для энергии $\mathcal{E}$ явное выражение $\mathcal{E} = \frac{3}{2} N k_B T$, можно получить искомые α и β :

$$\alpha = \ln k_B T, \quad \beta = \frac{1}{k_B T}. \quad (3.7)$$

Можно сказать, что α — коэффициент «информационности», а β — коэффициент «хаотичности» распределения $W(\mathcal{E})$.

3.4.3 Определение информации для дискретных систем связи

Система связи дискретна, если в ней для передачи информации применяются дискретные сигналы.

В указанной дискретной системе используются следующие основные понятия:

$A_{\text{ист}} = \{x_i\}_1^I$ — алфавит источника (передатчика) — множество алфавитных символов (сообщений), которые может передать источник в канал связи;

$A_{\text{прием}} = \{y_j\}_1^J$ — алфавит приемника — множество алфавитных символов (сообщений), которые могут быть приняты приемником;

$\bar{P} = \{p_i\}_1^I$ — вероятности того, что источник генерирует определенные сообщения из алфавита $A_{\text{ист}}$ и посылает их в канал связи; p_i — вероятность того, что источник передал в канал сообщение x_i ; вероятности p_i образуют полный набор, $\sum_{i=1}^I p_i = 1$;

$\bar{Q} = \{q_j\}_1^J$ — вероятности того, что приемник принимает определенные сообщения из алфавита $A_{\text{прием}}$; q_j — вероятность того, что приемник принял сообщение y_j ; при этом, как и выше, $\sum_{j=1}^J q_j = 1$;

$\mathbb{P} = [p_{ij}]_{I \times J}$ — матрица вероятностей, в которой p_{ij} — вероятность, что переданное сообщение x_i принято как сообщение y_j ;

$(\forall i) \left[\sum_{j=1}^J p_{ij} = 1 \right]$ — условие «каждое сообщение обязательно будет принято, пусть в искаженном виде».

Вероятности $\bar{P}$, $\bar{Q}$ — характеристики источника и приемника; матрица вероятностей $\mathbb{P}$ характеризует качество канала передачи.

Алфавиты $A_{\text{ист}}$, $A_{\text{прием}}$ могут находиться в различных отношениях. В идеальном случае (отсутствие помех в канале передачи)

$$A_{\text{ист}} = A_{\text{прием}}.$$

Если в канале есть помехи и/или приемник и передатчик дают сбой, то множества $A_{\text{ист}}$ и $A_{\text{прием}}$ имеют лишь общую область пересечения:

$$A_{\text{ист}} \cap A_{\text{прием}} \neq \emptyset,$$

где $\emptyset$ — пустое множество.

В дискретных системах помимо безусловных вероятностей $\bar{P}$, $\bar{Q}$ и $\bar{P}$ фигурируют условные вероятности:

$p(i|j)$ — вероятность, что передано сообщение i при условии, что принято сообщение j ;

$p(j|i)$ — вероятность принять сообщение j при условии, что передано сообщение i ;

$p(A|B)$ — вероятность, что принято (передано) по меньшей мере одно сообщение из множества A при условии, что передано (принято) по крайней мере одно сообщение из множества B ;

при этом по определению

$$p(i|j) = p_{ij}/p_j, \quad p(j|i) = p_{ij}/p_i.$$

Теперь можно переходить к строгим определениям энтропии и информации.

3.4.4 Безусловная энтропия

Обозначим для краткости $X = A_{\text{ист}}$, $Y = A_{\text{прием}}$.

Безусловная энтропия источника:

$$H_{\text{ист}} = H(X) = - \sum_{i=1}^I p_i \log p_i = \sum_{i=1}^I p_i \log \left(\frac{1}{p_i} \right) \quad (3.8)$$

(здесь и далее в этом разделе $\log$ обозначает двоичный логарифм). Смысл: $H(X)$ — средняя информация, приходящаяся на одно сообщение источника.

Безусловная энтропия приемника:

$$H_{\text{прием}} = H(Y) = - \sum_{j=1}^J p_j \log p_j. \quad (3.9)$$

Ее смысл аналогичен: средняя информация, приходящаяся на одно принятое сообщение безотносительно к тому, какое сообщение x передано в канал.

Существует безусловная совместная энтропия

$$H(X \cap Y) = - \sum_{i=1}^I \sum_{j=1}^J p_{ij} \log p_{ij}. \quad (3.10)$$

Смысл: среднее количество информации, приходящееся на любую пару сообщений (x, y) , $x \in X$, $y \in Y$.

Пример 1. Сообщение s_1 , либо s_2 передается двумя импульсами длительности Δt , которые располагаются в k -й и l -й позициях кадра длиной T (рис. 3.4).

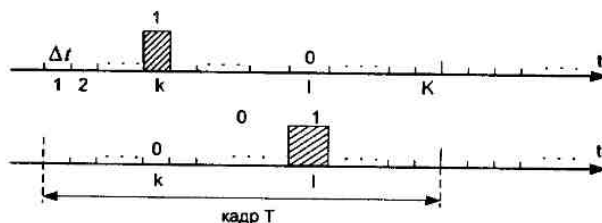


Рис. 3.4. Схема передачи сообщений $s_1 = 10$, $s_2 = 01$ в кадре

В открытом оптическом канале возможны помехи: поглощение импульса на атмосферных неоднородностях, $1 \rightarrow 0$, возникновение импульса в k -й или l -й позициях — случайная засветка, $0 \rightarrow 1$. Опыты говорят, что вероятность поглощения $p(1 \rightarrow 0)$ на несколько порядков превышает вероятность засветки $p(0 \rightarrow 1)$,

$$p(1 \rightarrow 0) \gg p(0 \rightarrow 1),$$

так что последней можно пренебречь.

Алфавит источника $X = \{01, 10\}$.

Алфавит приемника $Y = \{00, 01, 10, 11\}$.

Требуется найти безусловную совместную энтропию.

Решение. Составляем матрицу вероятностей $\mathcal{P}$ (табл. 3.1).

Таблица 3.1. Матрица вероятностей $\mathcal{P}$

Сообщения на выходе передатчика, X	Сообщения на входе приемника, Y			
	y_1	y_2	y_3	y_4
	00	01	10	11
$x_1 = 01$	α	$1 - \alpha$	0	0
$x_2 = 10$	β	0	$1 - \beta$	0

В табл. 3.1 α — вероятность искажения импульса в l -й позиции кадра, β — в k -й позиции. Величины $1 - \alpha$, $1 - \beta$ соответственно надежности передачи в указанных позициях.

По формуле (3.10) имеем

$$\begin{aligned} H(X \cap Y) &= -[\alpha \log \alpha + (1 - \alpha) \log(1 - \alpha) + 0 + 0 + \beta \log \beta + \\ &\quad + (1 - \beta) \log(1 - \beta) + 0 + 0] = \\ &= -1,4427[\alpha \ln \alpha + (1 - \alpha) \ln(1 - \alpha) + \beta \ln \beta + (1 - \beta) \ln(1 - \beta)]. \end{aligned}$$

Полагая, что внутрикадровые помехи не зависят от номера позиции, имеем $\alpha = \beta$ и выражение

$$H(X \cap Y) = -2,8854[\alpha \ln \alpha + (1 - \alpha) \ln(1 - \alpha)],$$

которое при малых $\alpha \sim 0,1$ приближенно равно

$$\begin{aligned} H(X \cap Y) = \tilde{H}(\alpha) &\approx -2,90 [\alpha \ln \alpha - (1 - \alpha) \alpha] \approx \\ &\approx -2,90 [\alpha \ln \alpha - \alpha] \approx -2,90 \alpha (\ln \alpha - 1). \end{aligned}$$

Получаем численную оценку $\tilde{H}(\alpha)$ при $\alpha = 0,1$:

$$\tilde{H}(0,1) \approx -2,90 \cdot 0,1 \cdot (\ln 0,1 - 1) \approx -2,90 \cdot 0,1 \cdot (-3,3) \approx 1 \text{ [бит]}.$$

$$\tilde{H}(0,01) \approx -2,90 \cdot 0,01 \cdot (-5,6) \approx 0,1624 \approx 0,2 \text{ [бит]}.$$

$$\tilde{H}(0,01) \approx -2,90 \cdot 0,001 \cdot (-7,9) \approx 0,023 \approx 0,02 \text{ [бит]}.$$

С убыванием вероятности ошибки α энтропия падает, и довольно быстро падает: при уменьшении α на порядок сначала H падает в пять раз, а потом, при следующем падении α , — тоже на порядок. Правильно: повышение надежности канала снижает неопределенность передачи.

Энтропия $H(X \cap Y)$ максимальна при $\alpha = \beta = 0,5$ и равна $H_{\max} = 2$ бита.

3.4.5 Условная энтропия

Наиболее важны условные энтропии $H(Y|x)$, $H(X|y)$, $H(X|Y)$ и $H(Y|X)$. Условная энтропия приемника относительно выделенного сообщения x источника

$$\begin{aligned} H(Y|x_i) &= - \sum_{j=1}^J p(y_j|x_i) \log p(y_j|x_i) = - \sum_{j=1}^J \frac{p_{ij}}{p_i} \cdot \log \left(\frac{p_{ij}}{p_i} \right) = \\ &= - \frac{1}{p_i} \sum_{j=1}^J p_{ij} (\log p_{ij} - \log p_i) = - \frac{\sum_{j=1}^J H(x_i, y_j)}{p_i} + \frac{\log p_i}{p_i} = \quad (3.11) \\ &= \frac{\log p_i - H(x_i)}{p_i}, \end{aligned}$$

где

$$H(x_i, y_j) = p_{ij} \cdot \log p_{ij} \quad \text{— безусловная энтропия пары } (x_i, y_j),$$

$$H(x_i) = \sum_{j=1}^J H(x_i, y_j) \quad \text{— усредненная энтропия переданного сообщения } x_i.$$

Смысл: среднее количество информации, приходящееся на каждое принятое сообщение, при условии, что источник выдает только сообщение x_i .

Пример 2. Пусть в примере 1 источник постоянно выдает сообщение $x_1 = 01$ с априорной вероятностью $p_1 = 0,5$. Требуется по табл. 3.1 определить условную энтропию приемника $H(Y|01)$.

Решение. По формуле (3.11) вычисляем:

$$\begin{aligned} H(Y|01) &= (\log \frac{1}{2} - \sum_{j=1}^4 p_{1j} \log p_{1j}) / 0,5 = \\ &= -[-1 - \alpha \log \alpha - (1 - \alpha) \log(1 - \alpha) - 0 - 0] / 0,5 = \\ &= -2(-1 - \frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2}) = 0. \end{aligned}$$

Получается, что сообщение 01 не несет в данном случае никакой информации.

Условная энтропия источника определяется сходным образом:

$$\begin{aligned} H(X|y_j) &= - \sum_{i=1}^I p(x_i|y_j) \log p(x_i|y_j) = - \sum_{i=1}^I \frac{p_{ij}}{q_j} \cdot \log \left(\frac{p_{ij}}{q_j} \right) = \\ &= - \frac{1}{q_j} \sum_{i=1}^I p_{ij} (\log p_{ij} - \log q_j) = - \frac{1}{q_j} [H(y_j) - \log q_j] = \\ &= \frac{\log q_j - H(y_j)}{q_j}, \end{aligned} \quad (3.12)$$

где $H(y_i) = \sum_{i=1}^I p_{ij} \cdot \log p_{ij}$ — усредненная энтропия принятого сообщения y_j .

Смысл: среднее количество информации, приходящееся на каждое переданное сообщение, при условии, что принято сообщение y_j , вероятность принять которое q_j .

Пример 3. Для той же табл. 3.1 (пример 1) найти условную энтропию $H(X|01)$.

Решение. По формуле (3.12) получаем:

$$\begin{aligned} H(X|01) &= [\log q_2 - \sum_{i=1}^2 p_{i2} \log p_{i2}] / q_2 = [\log q_2 - (1 - \alpha) \log(1 - \alpha) + 0] / q_2 = \\ &= \frac{1}{q_2} [\log q_2 - (1 - \alpha) \log(1 - \alpha)]. \end{aligned}$$

Для примера положим $q_2 = \frac{1}{4} \cdot \alpha = 0,01$. Тогда:

$$H(X|01) = 4 [2 - 0,99 \cdot \log 0,99] \approx 4 [2 + 0,014] \approx 8,06 \approx 8 \text{ [бит]}.$$

Это достаточно большая неопределенность. Можно сказать, что любое из сообщений X несет большую информацию, если на приемном конце возникает сообщение $Y = 01$.

Следующая энтропия $H(X|Y)$ носит название условной энтропии ненадежности канала:

$$\begin{aligned} H(X|Y) &= - \sum_{j=1}^J p(y_j) \cdot H(X|y_j) = - \sum_{j=1}^J q_j \cdot \frac{1}{q_j} [\log q_j - H(y_j)] = \\ &= - \sum_{j=1}^J [\log q_j - H(y_j)], \end{aligned} \quad (3.13)$$

где $H(y_j) = \sum_i p_{ij} \log p_{ij}$.

Энтропия $H(X|Y)$ — это количество информации, полученное усреднением информации, содержащейся в переданных сообщениях с учетом частоты их появления на входе приемника.

Название $H(X|Y)$ можно объяснить тем, что апостериорные вероятности $\{q_j\}_1^J$ приема сообщений зависят от процессов, происходящих в первую очередь в канале связи. Если бы канал не вносил искажений, то I было бы равно J и априорные вероятности p_i генерации сообщений совпадали бы с апостериорными вероятностями q_j приема этих же сообщений.

«Условная энтропия шума» $H(Y|X)$ определяется следующим образом:

$$\begin{aligned} H(Y|X) &= - \sum_{i=1}^I p(x_i) \cdot H(Y|x_i) = - \sum_{i=1}^I p_i \cdot \frac{1}{p_i} [\log p_i - H(x_i)] = \\ &= - \sum_{i=1}^I [\log p_i - H(x_i)], \end{aligned} \quad (3.14)$$

где $H(x_i) = \sum_j p_{ij} \log p_{ij}$.

$H(Y|X)$ — это количество информации, усредненное по всем полученным сообщениям источника с учетом частоты их передачи в канал.

Пример 4. Оценим условную энтропию ненадежности канала примера 1. Пусть $p_1 = p_2 = \frac{1}{2}$, т.е. сообщения $s_1 = 01$ и $s_2 = 10$ появляются в канале с равной вероятностью, и $q_1 = q_2 = q_3 = q_4 = \frac{1}{4}$ (с равной вероятностью принимаются все сообщения 00, 01, 10, 11).

$$\begin{aligned} H(X|Y) &= - \sum_{j=1}^4 [\log \frac{1}{4} - p_{1j} \cdot \log p_{1j} - p_{2j} \cdot \log p_{2j}] = \\ &= -4 \log \frac{1}{4} + \sum_{j=1}^4 (p_{1j} \cdot \log p_{1j} + p_{2j} \cdot \log p_{2j}) = \\ &= 8 + \alpha \cdot \log \alpha + (1 - \alpha \log \alpha) + 0 + 0 + \beta \log \beta + 0 + (1 - \beta) \log (1 - \beta) = \\ &= 8 + f(\alpha) + f(\beta). \end{aligned}$$

Очевидно

$$\max_{\alpha \in [0;1]} f(\alpha) = \max_{\beta \in [0;1]} f(\beta) = 0, \quad \text{а} \quad \min_{\alpha \in [0;1]} f(\alpha) = \min_{\beta \in [0;1]} f(\beta) = -1.$$

Отсюда следует

$$6 \leq H(X|Y) \leq 8 \text{ [бит]}.$$

Пример 5. Сравним полученную энтропию с условной энтропией шума $H(Y|X)$ при тех же $\bar{P} = \{p_1, p_2\}$ и $\bar{Q} = \{q_1, q_2, q_3, q_4\}$.

Имеем:

$$\begin{aligned} H(Y|X) &= - \sum_{i=1}^2 [\log p_i - \sum_{j=1}^4 p_{ij} \log p_{ij}] = \\ &= - [\log p_1 - \sum_{j=1}^4 p_{1j} \log p_{1j} + \log p_2 - \sum_{j=1}^4 p_{2j} \log p_{2j}] = \\ &= - [-2 - \alpha \log \alpha - (1 - \alpha) \log (1 - \alpha) + 0 + 0 - \beta \log \beta + \\ &\quad + 0 - (1 - \beta) \log (1 - \beta) + 0] = \\ &= 2 + \underbrace{\alpha \log \alpha + (1 - \alpha) \log (1 - \alpha)}_{-1 \leq f(\alpha) \leq 0} + \underbrace{\beta \log \beta + (1 - \beta) \log (1 - \beta)}_{-1 \leq f(\beta) \leq 0}. \end{aligned}$$

Как и в примере 3, получаем:

$$0 \leq H(Y|X) \leq 2.$$

При заданных априорных $\bar{P}$ и апостериорных $\bar{Q}$ вероятностях передачи и приема и при любых вероятностях сбоев α и β энтропия шума заведомо меньше энтропии ненадежности канала. Можно сказать, что на качество приема в большей степени влияет способ организации связи, чем помехи в канале. Необходимо всегда обращать внимание на процессы во всем тракте.

Любопытно сравнить полученные значения энтропий и на этой основе давать характеристику процессам передачи информации.

Полученные результаты можно свести в таблицу. Для алфавитов $X = \{01, 10\}$, $Y = \{00, 01, 10, 11\}$ и матрицы табл. 3.1 имеем табл. 3.2.

Таблица 3.2. Значения энтропии и характеристика процессов передачи информации для алфавитов и матрицы табл. 3.1

Характеристики	Области значений
1. Априорные вероятности $\bar{P}$ (вероятности передачи)	$\bar{p} = \{\frac{1}{2}; \frac{1}{2}\}$
2. Апостериорные вероятности $\bar{Q}$ (вероятности приема)	$\bar{q} = \{\frac{1}{4}; \frac{1}{4}; \frac{1}{4}; \frac{1}{4}\}$
3. $H(Y 01)$	0 бит
4. $H(X 01)$	8 бит
5. $H(X Y)$	[6; 8] бит
6. $H(Y X)$	[0; 2] бит
7. Безусловная взаимная энтропия $H(X \cap Y) = H(Y \cap X)$	[0; 2] бит

Заметно, что в табл. 3.2 условные энтропии передатчика значительно выше, чем условные энтропии приемника. Это значит, что при разработке каналов связи главное внимание надо уделять согласованию характеристик передатчика с режимом поступления сообщений. Согласованию характеристик приемника с параметрами канала связи в данном случае можно уделить меньше внимания.

Грубое правило «чем выше энтропия, тем более хаотичен процесс передачи и ему следует уделять больше внимания», действует на практике весьма неплохо. Все признают, что начинать наводить порядок надо с того места, в котором порядка меньше (разборку хлама начинайте там, где этого хлама больше всего). Но не всегда понимают, что это полезное правило есть следствие теории информации.

3.4.6 Взаимная информация

Мы уже знаем, что информация — это уменьшение энтропии системы связи.

Пусть мы находимся на передающей стороне и $H(X)$ — энтропия системы до передачи некоторого сообщения x из алфавита X в канал.

Пусть теперь $H(X|Y)$ — условная (остаточная) энтропия системы после того как, на приемном конце принято некоторое сообщение y из алфавита Y . $H(X)$ —

неопределенность, приходящаяся в среднем на одно сообщение источника. Эта средняя неопределенность, которую устраняет любое переданное сообщение x , выбранное источником из всего множества возможных сообщений X . Пусть сообщение x , пройдя через канал, трансформировалось в сообщение y , «выбранное» приемником из набора Y . Прием сообщения $y \in Y$, понятно, снижает начальную неопределенность $H(X)$, присущую системе. Тогда информация получается равной

$$I(X|Y) = H(X) - H(X|Y). \quad (3.15)$$

Информация (3.15) равна степени снижения начальной неопределенности, присущей источнику.

Перенесемся теперь на приемный конец. Пусть $H(Y)$ — энтропия системы, определенная по статистике Q принятых сообщений $y, y \in Y$. Приемник «не знает», какие при этом сообщения x передавались источником в канал. Приемник «видит» только факты: сообщение y_1 приходило с частотой q_1 , сообщение y_2 — с частотой $q_2, \dots$, сообщение y_j — с частотой q_j . Неопределенность этих фактов — $H(Y)$.

Допустим, проделан опыт, в котором фиксировались как переданные x , так и принятые сообщения y , и получена условная энтропия $H(Y|X)$, т. е. средняя информация, приходящаяся на сеанс связи $(x_i, y_j), i \in \{I\}, j \in \{J\}$, с учетом априорных вероятностей $\bar{P}$ появления сообщений X , причем об этих вероятностях «знает» только источник, приемнику они неизвестны.

Информация, полученная на приемном конце, будет равна:

$$I(Y|X) = H(Y) - H(Y|X). \quad (3.16)$$

Удивительно, но

$$I(X|Y) = I(Y|X). \quad (3.17)$$

Этот факт легко доказать с помощью введенных выше определений.

Теорема. Информации $I(Y|X)$ и $I(X|Y)$ совпадают и могут служить характеристикой системы связи в целом.

Докажем, что

$$H(X) - H(X|Y) = H(Y) - H(Y|X).$$

Доказательство.

$$I(X|Y) = \overline{\log(p_{i|j}/p_i)} = \log \frac{p_{ij}}{p_i \cdot q_j}$$

$$I(Y|X) = \overline{\log(p_{j|i}/q_j)} = \log \frac{p_{ij}}{p_i \cdot q_j}$$

Правые части равны, значит равны и левые:

$$I(X|Y) = I(Y|X) = I(X \cap Y).$$

Можно доказать иначе: из равенства $H(X \cap Y) = H(Y \cap X)$.

Следствие:

$$I(X \cap Y) = H(X) + H(Y) - H(X \cup Y).$$

Ввиду равенства (тождества) (3.17) информация

$$\mathbb{I}(X \cap Y) = \mathbb{I}(X|Y) = \mathbb{I}(Y|X)$$

может быть названа «взаимной». Она характеризует состояние (качество) системы в целом безотносительно к тому, где у системы вход, где выход. Система связи характеризуется независимо от того, где мы находимся: в канале, на приемном конце или на передающем. Алфавиты X, Y , вероятности $\bar{P}, \bar{Q}$ и $\bar{P}$, полностью характеризуют качество сети связи с точки зрения передачи информации.

Существуют легко доказываемые свойства энтропии и информации, весьма полезные при оценивании качества реальных каналов. Вот наиболее известные из них.

3.4.7 Свойства энтропии и информации

$$H(X) \geq H(X|Y), \quad (3.18)$$

$$H(Y) \geq H(Y|X); \quad (3.19)$$

$$H(X \cap Y) = H(X) + H(Y|X), \quad (3.20)$$

$$H(Y \cap X) = H(Y) + H(X|Y), \quad (3.21)$$

$$H(X \cap Y) = H(Y \cap X). \quad (3.22)$$

Из аксиомы теории вероятностей

$$P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$$

следует:

$$\begin{aligned} \mathbb{I}(X \cap Y) &= \mathbb{I}(Y \cap X) = H(X) - H(X|Y) = H(Y) - H(Y|X) = \\ &= H(X) + H(Y) - H(X \cap Y), \end{aligned} \quad (3.23)$$

где $H(X)$ — априорная неопределенность источника;

$H(Y)$ — априорная неопределенность приемника;

$H(X \cap Y)$ — неопределенность взаимосвязи «вход-выход».

В развитой математической теории энтропия рассматривается как мера некоторых множеств, на чем, как и в теории вероятностей, строятся доказательства многих важных свойств каналов связи.

Основное свойство: пусть $H(X), H(Y)$ — безусловные энтропии алфавитов X и Y ; если алфавиты не пересекаются, $X \cap Y = \emptyset$, то энтропия объединения алфавитов $X \cup Y$ должна быть равна, очевидно, сумме энтропий (свойство аддитивности):

$$H(X \cup Y) = H(X) + H(Y). \quad (3.24)$$

Энтропия разности алфавитов, по аналогии с теорией вероятностей, должна быть равна условной энтропии:

$$H(X \setminus Y) = H(X|Y), \quad (3.25)$$

$$H(Y \setminus X) = H(Y|X). \quad (3.26)$$

Из (3.24)–(3.26) следует большое количество полезных соотношений, очень удобных для получения оценок разнообразных информаций, возникающих в сложных сетях связи с многими алфавитами сообщений $X, Y, Z, \dots$

Пожалуй, наиболее интересным и важным свойством является обобщенная мультипликативность: энтропия пересечения алфавитов X и Y равна сумме безусловной и условной энтропий:

$$H(X \cap Y) = H[X \cup (X \setminus Y)] = H(X) + H(X \setminus Y) = H(X) + H(X|Y). \quad (3.27)$$

С помощью последнего соотношения легко доказывается формула (3.17) и множество прочих полезных формул.

Для системы связи без потерь (при отсутствии или пренебрежимой малости потерь в открытом канале) количество передаваемой информации, очевидно, равно энтропии источника:

$$\mathbb{I}(X, Y) = H(X) = H(Y) = H(X, Y), \quad (3.28)$$

$$H(Y|X) = H(X|Y) = 0. \quad (3.29)$$

Количество передаваемой информации $\mathbb{I}(X, Y) = \mathbb{I}(Y, X)$ равно нулю, когда вход и выход канала связи независимы:

$$\begin{aligned} H(X|Y) &= H(X), \\ H(Y|X) &= H(Y), \\ \mathbb{I}(X, Y) &= \mathbb{I}(Y, X) = 0. \end{aligned} \quad (3.30)$$

3.4.8 Пропускная способность дискретного канала

Это наиболее важный для инженеров-практиков показатель качества канала, впервые как понятие введенный К. Шенноном в 1948 году.

Если среди источников существует такой, при работе которого по каналу передается максимальное количество информации $C = \max \mathbb{I}(X, Y)$, то величина C называется пропускной способностью, а источник считается согласованным с каналом.

$$C = \max_{p_1, p_2, \dots, p_I} \mathbb{I}(X, Y; \bar{P}, \bar{Q}, \textcircled{P}). \quad (3.31)$$

Для оценивания пропускной способности канала необходимо максимизировать информацию $\mathbb{I}$ как функцию от I аргументов $\bar{P}$ при фиксированных алфавитах X, Y и заданных апостериорных вероятностях $\bar{Q}$ и матрице $\textcircled{P}$.

Существуют задачи, в которых максимизацию $(2 + I + J + I \cdot J)$ -аргументной функции

$$\mathbb{I}(X, Y; p_1, p_2, \dots, p_I, q_1, q_2, \dots, q_J, \{p_{ij}\}_{I \times J})$$

нужно проводить не только по переменным источника $p_i, i = 1, 2, \dots, I$, но и по апостериорным вероятностям $\bar{Q}$ и по параметрам канала $\textcircled{P}$. Максимизация функции $\mathbb{I}$ по априорным вероятностям — лишь наиболее известная традиционная задача, впрочем, весьма непростая и решаемая методами условной экстремизации (метод множителей Лагранжа).

Пример 6. Для простейшего случая $X = \{0; 1\} = Y$ и вероятности искажения $p(0 \rightarrow 1) = p(1 \rightarrow 0) = p$ пропускная способность составит

$$C = 1 + p \log p + (1 - p) \log(1 - p),$$

при этом энтропия источника составляет 1 бит/сообщение.

3.4.9 Скорость передачи

Средняя скорость передачи информации v по каналу связи при передаче одного сообщения (или последовательной передаче N сообщений) определяется как

$$v = \frac{1}{T} \cdot \mathbb{I}(X, Y), \quad (3.32)$$

где T — длина кадра, в течение которого передается сообщение из алфавита X , а $\mathbb{I}(X, Y)$ — взаимная информация системы связи. Если источник согласован с каналом, то

$$v = v_{\max} = C/T, \quad (3.33)$$

где C — пропускная способность последовательного канала.

3.4.10 Аналоговые источники

Аналоговая (непрерывная) передача осуществляется, если источник передает непрерывный сигнал S с энергетической характеристикой x , плотность распределения вероятностей которой $w(x)$,

$$\int_X w(x) dx = 1,$$

где X — область значений x . Тогда энтропия сигнала может быть определена как

$$H(X) = E[-\log w(x)], \quad (3.34)$$

где E — символ математического ожидания,

$$E[-\log w(x)] = - \int_{-\infty}^{\infty} w(x) \log w(x) dx.$$

Интегралы типа (3.34) существуют не для всех плотностей $w(x)$ и зависят от выбранной шкалы измерений, однако часто приносят пользу в теоретических выкладках. По аналогии с дискретным случаем существуют энтропии

$$H(Y|X) = -E[\log w(y|x)] = - \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} w(x, y) \log \frac{w(x, y)}{w_1(x)} dx, \quad (3.35)$$

$$H(X|Y) = -E[\log w(x|y)] = - \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} w(x, y) \log \frac{w(x, y)}{w_2(y)} dy, \quad (3.36)$$

$$H(X, Y) = -E[\log w(x, y)] = - \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} w(x, y) \log w(x, y) dx. \quad (3.37)$$

Информация $I(X; Y)$, содержащаяся в непрерывном сигнале S с областями значений X и Y на передающем и приемном концах, соответственно, составит

$$I(X, Y) = \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} w(x, y) \log \frac{w(x, y)}{w_1(x) w_2(y)} dx. \quad (3.38)$$

3.4.11 Максимизация энтропии

Для аналоговой системы связи всегда полезно найти маргинальное распределение вероятностей сигналов X источника $w_1(x)$, которое максимизировало бы его энтропию. Хорошо известны 3 случая (К. Шеннон):

1) источник с

$$w_1(x) = \frac{1}{x_{\max} - x_{\min}}, \quad x \in [x_{\min}; x_{\max}] \quad (3.39)$$

максимизирует энтропию

$$H_{\max} = \ln(b - a) \quad (3.40)$$

сигналов, амплитуда, частота или другая характеристика x определена на отрезке $[x_{\min}; x_{\max}]$;

2) источник сигналов, определенных на $[0; \infty]$ и имеющих фиксированное среднее $E[x] = \bar{x}$, максимизирует энтропию, когда

$$w_1(x) = \frac{1}{x} e^{-x/\bar{x}}, \quad (3.41)$$

где $\bar{x}$ — математическое ожидание характеристики x ;

$$H_{\max} = \ln(\bar{x}e); \quad (3.42)$$

3) источник сигналов, определенных на $(-\infty; +\infty)$ и имеющих фиксированное среднее $E[x] = \bar{x}$ и дисперсию $E[x - E(x)]^2 = \sigma^2$, максимизирует энтропию при нормальном распределении

$$w_1(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x - \bar{x})^2}{2\sigma^2} \right], \quad (3.43)$$

при этом

$$H_{\max} = \ln \sqrt{2\pi e} \sigma \quad (3.44)$$

не зависит от математического ожидания $\bar{x}$ характеристики x , а лишь от ее дисперсии.

Гауссовы каналы. Для аналоговых сигналов получен ряд практически важных результатов. Так, если вход x и выход y открытого канала без помех подвержены случайным нормальным колебаниям и коррелированы (т. е. в канал поступает сигнал с колоколообразным спектром или поступающий детерминированный сигнал подвергается воздействию нормальных помех с интенсивностью σ_x^2 и на входе приемника имеются помехи с интенсивностью σ_y^2 либо их вносит сам

канал), то

$$w(x, y) = \frac{1}{2\sigma_x\sigma_y\sqrt{1-\rho^2}} \times \\ \times \exp \left[-\frac{1}{\sqrt{1-\rho^2}} \left(\frac{(x-\bar{x})^2}{\sigma_x^2} - 2\rho \frac{(x-\bar{x})(y-\bar{y})}{\sigma_x\sigma_y} + \frac{(y-\bar{y})^2}{\sigma_y^2} \right) \right] \quad (3.45)$$

и

$$\mathbb{I}(X, Y) = -\frac{1}{2} \ln(1 - \rho^2), \quad (3.46)$$

где ρ — коэффициент корреляции входа и выхода, $|\rho| \leq 1$.

Весьма важный результат получен для аналогового сигнала с аддитивной помехой. Пусть x — сигнал на входе канала, а ε — помехи в канале, и пусть сигнал y на выходе канала и, стало быть, на входе приемника определяется как сумма

$$y = x + \varepsilon. \quad (3.47)$$

(Для простоты можно принять y, x, ε амплитудами сигналов.) Количество передаваемой информации записывается в виде

$$\mathbb{I}(X, Y) = H(Y) + \int_{-\infty}^{\infty} W(\varepsilon) \log W(\varepsilon) d\varepsilon, \quad (3.48)$$

где $W(\varepsilon)$ — интегральный закон распределения помех ε .

Для случая, когда сигнал x на входе подчинен нормальному закону распределения с нулевым средним и дисперсией σ_x^2 , а помехи ε также нормальны с $\bar{\varepsilon} = 0$ и интенсивностью (дисперсией) σ_ε^2 , пропускная способность канала составит

$$C = \max \mathbb{I}(X, Y) = \frac{1}{2} \log \left[1 + \left(\frac{\sigma_x}{\sigma_\varepsilon} \right)^2 \right] = \frac{1}{2} \log [1 + P_c/P_{\text{ш}}], \quad (3.49)$$

где $\sigma_x^2/\sigma_\varepsilon^2$ — известное отношение энергий (мощностей) сигнал/шум.

Когда передача идет с помощью N сигналов (вектор N некоррелированных сигналов), а помехи независимы и имеют одинаковую интенсивность по каждой компоненте, то пропускная способность такого канала составит

$$C = \frac{N}{2} \log (1 + P_c/P_{\text{ш}}), \quad (3.50)$$

где $P_c/P_{\text{ш}}$ — отношение сигнал/шум по одной компоненте вектора сигналов.

3.5 Теорема отсчетов и ее использование для выбора режимов функционирования канала связи

3.5.1 Теорема отсчетов

Под этим термином в теории сигналов и теории информации фигурирует достаточно широкое семейство утверждений, которым по традиции придана фор-

ма теорем. Эти теоремы определяют условия, при которых аналоговый сигнал $s(t)$ можно точно восстановить по дискретным отсчетам $s_1 = s(t_1)$, $s_2 = s(t_2)$, ..., $s_k = s(t_k)$, ..., расположенным через определенные промежутки времени $\tau_1 = t_2 - t_1$, $\tau_2 = t_3 - t_2$, Начиная с пионерской работы В.А. Котельникова 1933 года [49] (на Западе приоритет отдают Х. Найквисту) возникают разнообразные обобщения и толкования теоремы отсчетов, предпринимаются попытки применить ее в радиотехнических цепях.

Удивительно, но эта знаменитая теорема называется теоремой только в учебниках по радиотехнике и связи. Математики ее почти не замечают, в математических справочниках ее нет. Нет ее и в последнем издании математической энциклопедии.

Теоремы отсчета имеют дело с представлениями аналогового сигнала $s(t)$ в виде

$$s(t) \underset{t \in \{T\}}{\approx} \sum_k f_k(s) \varphi_k(t) = \psi(t), \quad (3.51)$$

где $\{\varphi_k\}$, $k = 1, 2, \dots$ — множество сигналов (функций времени), выбранных заранее независимо от вида $s(t)$, $\{T\}$ — множество моментов времени, на котором определены $s(t)$ и $\{\varphi_k(t)\}$, $k = 1, 2, \dots$

Знак $\approx$ указывает на то, что правая часть (3.51) не обязательно совпадает с $s(t)$ во всех точках, а лишь близка к $s(t)$ в соответствии с выбранной метрикой (расстоянием) $\rho[s(t), \psi(t)]$.

Наиболее часто встречается квадратичная метрика вида

$$\rho_{sq} = \frac{1}{2T} \int_{-T}^{+T} |s(t) - \psi(t)|^2 dt, \quad (3.52)$$

пропорциональная, как правило, разности энергий аппроксимируемого $s(t)$ и аппроксимирующего $\psi(t)$ сигналов.

В канале связи под сигналом $s(t)$ понимается огибающая амплитуды несущего сигнала $S(t)$ частоты порядка $f \sim 10^{14...15}$ Гц. Приблизительный вид сигнала в канале связи $s(t)$ на ограниченном отрезке времени $\{T\}$ показан на рис. 3.5 (жирная линия).

В роли функций $\{\varphi_k(t)\}$ обычно используются импульсы, смещенные относительно начала отсчета времени пропорционально k . Импульс $\varphi_k(t)$ называют интерполирующим, если

$$(\forall k) [\varphi_k(0) = 1], \quad (\forall l = 1, 2, \dots) [\varphi_k(l\tau) = 0], \quad (3.53)$$

где τ — некоторый конечный интервал времени, период отсчетов.

В роли функций $f_k(t)$ для удобства реализации в большинстве случаев выступают точечные отсчеты

$$f_k(t) = S(t_k) = s(k\tau), \quad k = 0, \pm 1, \pm 2, \dots, \quad (3.54)$$

а вид импульсов $\{\varphi_k(t)\}$ принимается одинаковым. Тогда

$$S(t) \approx \sum_k s(k\tau) \varphi(t - k\tau), \quad -\infty \leq -T \leq t < T \leq \infty. \quad (3.55)$$

Представление (3.55) дает, очевидно, точное равенство при $t = k\tau$, и, если $s(t)$ в промежутках между отсчетами изменяется не слишком быстро, погрешность

$$\rho \left(s(t); \sum_k s(k\tau) \cdot \varphi(t - k\tau) \right)$$

получается технически допустимой. Выражение (3.55) — наиболее слабая форма теоремы отсчетов.

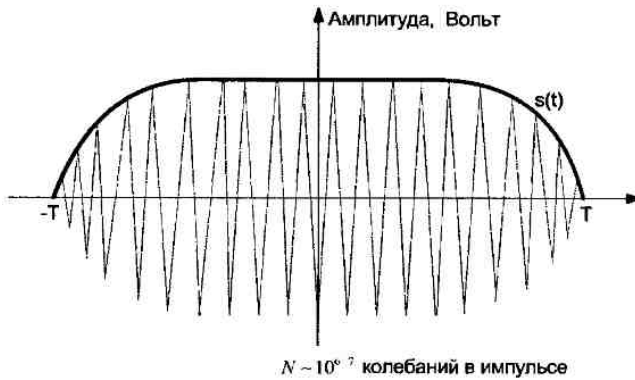


Рис. 3.5. Вид огибающей сигнала $s(t)$ в открытом атмосферном канале связи

3.5.2 Теорема отсчетов для сигналов с ограниченной полосой

Более сильное утверждение справедливо для сигналов с ограниченной полосой частот. Для них спектр

$$\tilde{S}(\omega) = 0 \quad (3.56)$$

вне интервала $|\omega| \leq \Omega$, где $\Omega = 2\pi F$ — максимально возможная круговая частота спектра. Внутри интервала $|\omega| \leq \Omega$ спектр $\tilde{S}(\omega)$ отличен от нуля по крайней мере на некотором подмножестве точек ω ненулевой меры (область, в точках которой $\tilde{S}(\omega) \neq 0$, имеет ненулевую длину). Сделанная оговорка имеет, скорее, формальный характер, ибо нет смысла рассматривать тождественно нулевой спектр, но спасает от многих математических проблем.

Спектр аналогового сигнала $S(t)$, как известно, определяется формулой

$$\tilde{S}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} s(t) e^{-i\omega t} dt; \quad (3.57)$$

при этом $\tilde{S}(\omega)$ интегрируется в квадрате на $[-\Omega, +\Omega]$, ибо предполагается, что сигнал

$$S(t) = \frac{1}{\sqrt{2\pi}} \int_{-\Omega}^{+\Omega} \tilde{S}(\omega) e^{-i\omega t} d\omega \quad (3.58)$$

обладает конечной энергией в области своего определения $-\infty < t < +\infty$ (теоретически бесконечной)

$$\int_{-\Omega}^{+\Omega} |\tilde{S}(\omega)|^2 d\omega < \infty. \quad (3.59)$$

3.5.3 Теорема 1 Котельникова. Сигнал $s(t)$ задан на бесконечной оси времени, $-\infty < t < +\infty$

Теорема 1. Для аналоговых сигналов $s(t)$ с ограниченной полосой $\tilde{s}(\omega)$ справедливо равенство

$$\begin{aligned} (\forall t \in (-\infty; \infty)) (\forall \tau > 0) \quad s(t) &= \sum_{k=-\infty}^{\infty} s(k\tau) \frac{\sin \frac{\pi}{\tau}(t - k\tau)}{\frac{\pi}{\tau}(t - k\tau)} = \\ &= \sum_{k=-\infty}^{\infty} s\left(\frac{k}{2F}\right) \frac{\sin 2\pi F\left(t - \frac{k}{2F}\right)}{2\pi F\left(t - \frac{k}{2F}\right)} = \sum_{k=-\infty}^{\infty} s\left(\frac{k}{2F}\right) \frac{\sin \pi(2Ft - k)}{\pi(2Ft - k)} = \\ &= \sum_{k=-\infty}^{\infty} s\left(\frac{k\pi}{\Omega}\right) \frac{\sin \left[\Omega\left(t - \frac{k\pi}{\Omega}\right)\right]}{\Omega\left(t - \frac{k\pi}{\Omega}\right)} = \sum_{k=-\infty}^{\infty} s\left(k\frac{\pi}{\Omega}\right) \frac{\sin(\Omega t - k\pi)}{\Omega t - k\pi}, \end{aligned} \quad (3.60)$$

где τ — интервал времени между двумя соседними равноотстоящими отсчетами сигнала $s(t)$,

$$\tau = \frac{1}{2F} = \frac{\pi}{\Omega}. \quad (3.61)$$

Частоту $2F$ иногда называют частотой Найквиста.

Представление сигнала $s(t)$ рядом (3.60) иллюстрировано на рис. 3.6. Сплошная линия $s(t)$ представляет собой сумму сплошного и пунктирных горбов $\varphi_k(t)$, умноженных на величину амплитуд $s(k\tau)$ в точках $k\tau$.

3.5.4 Физический смысл теоремы 1 Котельникова

Теорема (3.60), справедливая при условии (3.61), говорит о том, что операции дискретизации (цифрового преобразования) и восстановления сигнала взаимно обратны. Из разложения (3.60) следует, что для любого задаваемого t аналоговый сигнал $s(t)$ теоретически можно получить на приемном конце, если пропустить всю бесконечную последовательность $s(k\tau)$, $k = \dots, -2, -1, 0, +1, +2, \dots$, через идеальный фильтр нижних частот с частотой среза $\Omega = \pi/\tau$ и амплитудно-частотной характеристикой (АЧХ) $|K(i\omega)| = \tau$ в полосе пропускания. Однако идеального фильтра не существует и весьма проблематично пропустить через него (даже если бы он существовал) бесконечную последовательность отсчетов.

Поэтому теорема 1 Котельникова относится к гипотетической ситуации и характеризует лишь предельные возможности канала связи.

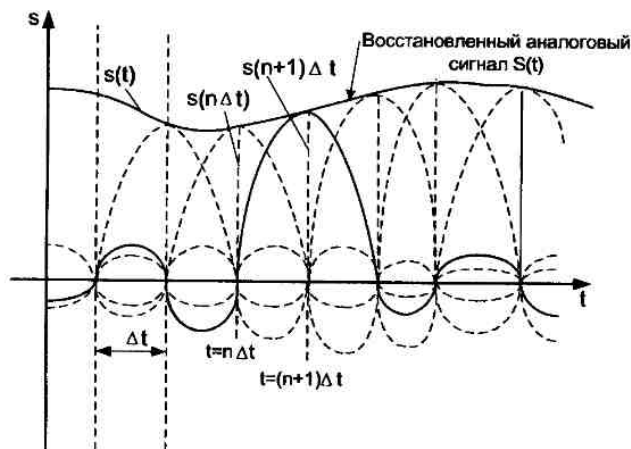


Рис. 3.6. Представление сигнала $s(t)$ рядом (3.60) (см. текст)

Если быть до конца точным, то выражение (3.60) следует понимать в предельном смысле:

$$(\forall t)(\forall \tau) \quad s(t) = \lim_{k \rightarrow \infty} \underbrace{\sum_{k=-K}^k s(k\tau) \cdot \frac{\sin \pi(\frac{t}{\tau} - k)}{\pi(\frac{t}{\tau} - k)}}_{\psi_k(t; \tau)}, \quad (3.62)$$

причем сходимость определяется в квадратичной метрике

$$\rho(s(t); \psi(t; \tau)) = \int_{t=-\infty}^{t=\infty} |s(t) - \psi(t)|^2 dt, \quad (3.63)$$

где

$$\psi(t; \tau) = \lim_{k \rightarrow \infty} \psi_k(t; \tau). \quad (3.64)$$

Доказательство теоремы Котельникова является весьма простым. Поскольку спектр $\tilde{s}(\omega)$ ограничен в частотной области отрезком $[-\Omega; +\Omega]$, его можно представить рядом Фурье с коэффициентами

$$C_k = \int_{-F}^{+F} \tilde{s}(2\pi f) e^{i2\pi f \cdot \frac{k}{2F}} df, \quad k = 0, \pm 1, \pm 2, \dots$$

Сопоставив последнее выражение с формулой (3.58), убеждаемся, что коэффициенты Фурье совпадают с отсчетами сигнала $s(t)$ в моменты $t_k = k/2F$,

$$C_k = s(k/2F) = s(k\tau).$$

Подставим в (3.58) представление спектра $\tilde{s}(2\pi f)$ в виде ряда Фурье, получим

$$s(t) = \int_{-F}^{+F} \left[\sum_{k=-\infty}^{\infty} C_k e^{-i2\pi f \cdot \frac{k}{2F}} \right] e^{i2\pi f \cdot t} df.$$

Проинтегрировав последнее выражение и заменив C_k на $s(\frac{k}{2F})$, получим одно из выражений (3.60):

$$s(t) = \sum_{k=-\infty}^{\infty} s\left(\frac{k}{2F}\right) \frac{\sin \pi(2Ft - k)}{\pi(2Ft - k)}.$$

Теорема доказана.

В математическом анализе, точнее, в теории аппроксимации и теории ортогональных функций (многочленов), в частности полиномов Чебышева, только что использованный нами прием доказательства относится к стандартным. Тригонометрические функции $\sin(2\pi kt/T)$ и $\cos(2\pi kt/T)$, по которым ведется Фурье-анализ, — частный случай ортогональных функций, поэтому математики считают теорему Котельникова всего лишь частным случаем теоремы интерполяции в теории целых (разложимых в сходящийся ряд Тейлора на всей оси t или на всей комплексной плоскости) функций. Формула (3.60) в 1930–1935 годах появлялась в работах Уиттеккера, Ватсона, отечественного математика Гончарова и др. Но лишь Котельников впервые увидел в ней физический смысл и указал инженерные применения. Это замечательное достижение. Так, формулы сокращения размеров движущихся предметов вывели Лоренц и Пуанкаре, а Эйнштейн объяснил их, приложил формулы к практике. Через 40 лет атомная бомба убедительно подтвердила прозорливость Эйнштейна.

Теорема Котельникова в том виде, в каком сформулирована, может использоваться для инженерных надобностей следующим образом. С какого-то фиксированного момента t_0 на приемный конец канала связи начинают поступать отсчеты $s(kt)$. Точечный отсчет конечно же снова идеализация. Отсчеты реализуются в виде импульсов прямоугольной (треугольной, трапецевидной или иной аналитически заданной) формы. Длительность импульсов $\Delta\tau$ достаточно малая, чтобы в интервале τ успели закончиться все переходные процессы в приемнике. Так, используемая ныне стандартная аппаратура ОАК (открытого атмосферного канала) позволяет без труда формировать импульсы длительности

$$\Delta\tau \simeq 10 \dots 50 \text{ нс} = (1 \dots 5) \cdot 10^{-8} \text{ с}$$

с почти прямоугольной огибающей, надежно гарантируя интервал дискретизации не более

$$\tau \simeq 100 \dots 200 \text{ нс} = (1 \dots 2) \cdot 10^{-7} \text{ с}.$$

При этом возможна дискретизация сигналов с верхней частотой

$$F = \frac{1}{2\tau} \approx \frac{1}{(2 \dots 4) \cdot 10^{-7}} \approx (0,25 \dots 0,5) \cdot 10^7 \text{ Гц} \approx 30 \text{ МГц},$$

обеспечивающей скорость передачи порядка $30 \frac{\text{Мбит}}{\text{с}} \approx 3 \frac{\text{Мбайт}}{\text{с}}$.

Пусть на приемном конце вид сигнала $s(t)$ известен заранее. Импульсы $s(k\tau)$, поочередно поступаая на вход приемника, формируют текущую сумму ряда

$$\psi_k(t; \tau), \quad k = \left\lfloor \frac{t_0}{\tau} \right\rfloor, \left\lfloor \frac{t_0}{\tau} \right\rfloor + 1, \left\lfloor \frac{t_0}{\tau} \right\rfloor + 2, \dots,$$

для заранее выбранного $t = K\tau + \frac{\tau}{2}$ (или, что то же, для выбранного K). Величина t специально берется не кратной $K\tau$, чтобы оценить точность восстановления сигнала в точке, отличной от момента отсчета. Как только номер текущего отсчета совпадает с K , $k = K$, определяется значение суммы

$$\psi_K(t) = \psi_K(K\tau + \tau/2)$$

и по прошествии периода времени $\tau/2$ сравнивается с эталоном $s(t) = s(K\tau + \tau/2)$. Если погрешность восстановления

$$\Delta s = |s(t) - \psi_K(t)| \quad (3.65)$$

не превышает заданной максимальной величины,

$$\Delta s \leq \Delta s_{\max}, \quad (3.66)$$

то длина интервала дискретизации τ и количество отсчетов K достаточны для практически точного восстановления аналогового сигнала по его дискретным отсчетам.

3.5.5 Необходимое уточнение теоремы 1 Котельникова

Выше в наших рассуждениях появился произвольный начальный момент отсчета t_0 , в то время как теорема Котельникова сформулирована для фиксированного начального отсчета времени $t_0 = 0$. Что нужно изменить в формуле (3.60), чтобы учесть этот факт?

Из спектральной теории известно, что при сдвиге начала отсчета времени t_0 на фиксированную величину ΔT , $t_0 \rightarrow t_0 + \Delta T$,

$$|\Delta T| < \tau$$

спектр $\tilde{s}(\omega)$ умножается на $e^{\pm i\omega\Delta T}$, причем $|e^{\pm i\omega\Delta T}| = 1$; граничное условие (3.56), налагаемое на новый спектр $\tilde{s}(\omega) = \tilde{S}(\omega) \cdot e^{\pm i\omega\Delta T}$, остается справедливым, отчего формулировка теоремы не меняется, только в разложение (3.60) добавляется сдвиг начала отсчета ΔT :

$$s(t) = \sum_{k=-\infty}^{\infty} s(k\tau + \Delta T) \cdot \frac{\sin \frac{\pi}{\tau}(t - \Delta T - k\tau)}{\frac{\pi}{\tau}(t - \Delta T - k\tau)}. \quad (3.67)$$

Можно распространить теорему на сигналы конечной длительности, заданные на интервале $(-T; +T)$, $T < \infty$. Условие ограниченности спектра $\tilde{s}(\omega)$ и условие конечности области πmT определения сигнала $s(t)$ не могут быть выполнены одновременно. Если сигнал определен на конечном интервале, то его спектр существует и определен на бесконечном интервале частот, $-\infty < \omega < +\infty$. Верно и обратное: конечность области определения спектра $\tilde{s}(\omega)$ логически неизбежно влечет неограниченность области $(-T; +T)$. Ограниченные во времени сигналы

нельзя точно представить даже бесконечным рядом (3.60) для любого момента t , потому что их спектр безграничен.

Это утверждение верно для произвольных сигналов конечной энергии, разложимых по любой системе ортогональных функций: если функционал $\mathcal{F}[s(t)]$ существует только в ограниченной области изменения переменной, то для представления $s(t)$ потребуются бесконечное число отсчетов.

Тем не менее для сигналов конечной длительности $(-T; +T)$ всегда возможно выбрать граничную частоту $\Omega = 2\pi F$ таким образом, что за ее пределами влияние неучтенных спектральных составляющих будет мало. Сигнал $s(t)$ можно тогда представить достаточно точно с помощью

$$K = 2F \cdot 2T \quad (3.68)$$

отсчетов.

3.5.6 Можно ли переносить теорему на случайные сигналы?

Полученные результаты часто переносят на случайные сигналы $s(t)$, чья амплитуда или другие характеристики подвержены помехам. Полагают, что любая реализация сигнала представима $4F \cdot T$ отсчетами. Это не верно хотя бы потому, что указанные отсчеты в общем случае коррелированы. Если радиус корреляции Δt много меньше τ , то отсчеты можно считать статистически независимыми и применить к ним теорию информации для конечных алфавитов $\mathcal{A}_N$ и $\mathcal{B}_M$. При этом размерности алфавитов N и M будут совпадать с количеством K передаваемых в канале связи и принимаемых на его конце сигналов (отсчетов):

$$N = M = K. \quad (3.69)$$

Рассмотрим стационарный случайный сигнал $s(t)$. Пусть его амплитудный спектр имеет вид, показанный на рис. 3.7,

$$\tilde{s}(f) = \begin{cases} C = \text{const}, & \text{если } -F \leq f \leq F, \\ 0, & \text{в противном случае.} \end{cases} \quad (3.70)$$

Запишем выражение для корреляционной функции

$$R(\tau) = \int_{-F}^{+F} C \cdot e^{i2\pi f\tau} df = 2CF \cdot \frac{\sin 2\pi F\tau}{2\pi F\tau}, \quad (3.71)$$

где τ — как и выше, интервал времени между отсчетами. График $R(\tau)$ показан на рис. 3.8.

Функция $R(\tau)$ обращается в нуль при $\tau = 1/2F, 2/2F, 3/2F, \dots, k/2F$. Это значит, что отсчеты $s(k\tau)$, $k = 0, \pm 1, \pm 2, \dots$ не коррелированы между собой. Поскольку отсчеты в силу стационарности подчиняются одному и тому же закону распределения, они обладают одинаковой энтропией. Максимальное количество информации, содержащееся в сигнале длительностью $2T$ и с полосой частот

$(-F; +F)$, равно произведению энтропии одного отсчета

$$\int_{-\infty}^{\infty} \omega(s) \log \omega(s) ds = h$$

на число отсчетов

$$I(s) = 4F \cdot T \cdot h. \quad (3.72)$$

Рис. 3.7. Амплитудный спектр случайного сигнала $s(t)$

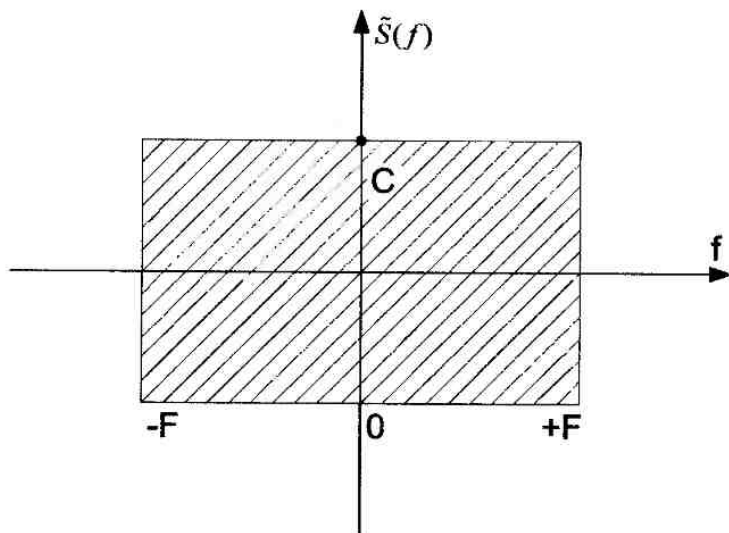
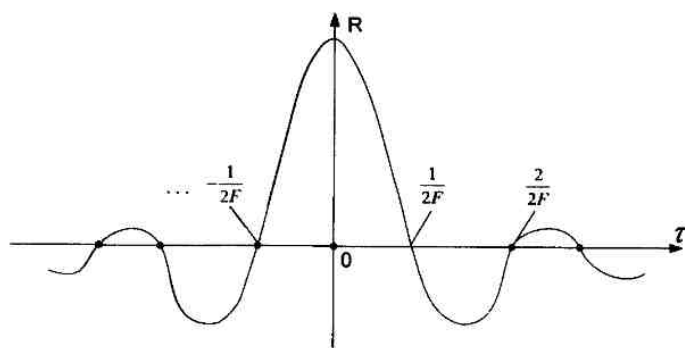


Рис. 3.8. График корреляционной функции $R(\tau)$ (см. текст)



Для случая нормального распределения относительно фиксированной амплитуды отсчетов-импульсов получим для информации, передаваемой по каналу связи, следующее полезное выражение:

$$I(Y|X) = 2F \cdot T \cdot h_{\sigma} = F \cdot T \cdot \log \left(\frac{\sigma^2}{\varepsilon^2} \right),$$

где σ — мощность сигнала $s(t)$, а ε — допустимая погрешность восстановления сигнала ($\varepsilon = \Delta s_{\max}$, см. выше (3.66)), X — передаваемый, Y — принимаемый сигналы. Скорость передачи составит

$$v = \frac{I}{T} = F \log \left(\frac{\sigma}{\varepsilon} \right)^2.$$

После всех этих подготовительных шагов перейдем к главному.

Запишем теорему отсчетов для стационарного случайного процесса с ограниченной полосой. Ограниченность полосы означает, что корреляционная функция на приемном конце $R_{yy}(f) = 0$ вне частотной полосы $|f| > \frac{1}{2\tau}$; при этом $y(t)$ — значения принимаемого сигнала Y .

Для любого f ввиду стационарности канала имеем:

$$R_{xx}(f) = \varphi(f) \sum_{l=-\infty}^{\infty} k_{yy}(l\tau) e^{i2\pi l\tau f} = \frac{\tilde{s}(f)}{\tau} \sum_{l=-\infty}^{\infty} R_{yy}(f - \frac{l}{\tau}) = R_{yy}(f).$$

Пусть для конкретности в канал передается АИМ сигнал

$$x(t) = \sum_{k=-\infty}^{\infty} x_k \cdot \varphi(t - k\tau), \quad (3.73)$$

где x_k — дискретные отсчеты амплитуды, являющиеся реализациями дискретного случайного процесса:

$$\varphi(t) = \frac{\sin(\pi t/\tau)}{\pi t/\tau},$$

при этом

$$\tilde{\varphi}(t) = \begin{cases} \tau, & \text{если } |f| \leq 1/2\tau, \\ 0, & \text{в противном случае.} \end{cases}$$

Расстояние между случайными сигналами $x(t)$ и $y(t)$ определяется часто используемым выражением, имеющим смысл средней энергии разности сигналов:

$$\rho(x, y) = E\{[x(t) - y(t)]^2\},$$

где E — символ математического ожидания. Раскрывая последнее выражение, получим

$$\begin{aligned} \rho(x, y) &= E[x^2] + E[y^2] - 2E[x(t) \cdot y(t)] = \\ &= k_{xx}(0) + k_{yy}(0) - 2 \sum_{k=-\infty}^{\infty} k_{yy}(t - k\tau) \cdot \varphi(t - k\tau), \end{aligned}$$

где $k_{xx}(t + \tau, t) = E[x(t + \tau) \cdot x(t)]$ — автокорреляционная функция случайного сигнала $x(t)$, $k_{yy}(t + \tau, t)$ — то же для $y(t)$, при этом ввиду (3.66) $2E[x \cdot y] = k_{yy}(0)$, так что

$$\rho(x, y) = 2[k_{yy}(0) - k_{yy}(0)] = 0.$$

Отсюда заключаем, что ограниченный по полосе процесс $y(t)$ на входе приемника может быть представлен рядом Котельникова

$$y(t) = \sum_{k=-\infty}^{\infty} y(k\tau) \frac{\sin[(\pi/\tau)(t - k\tau)]}{(\pi/\tau)(t - k\tau)}. \quad (3.74)$$

Представление (3.74) точно в том смысле, что средняя энергия разности сигналов $x(t)$ и $y(t)$ равна нулю.

3.5.7 Что получится, если в теореме Котельникова отсчеты брать чаще или реже, чем с периодом $\tau = 1/F$?

Чтобы дать ответ, изучим сначала некоторые спектральные характеристики разложения Котельникова (3.60) детерминированного (неслучайного) сигнала

$$s(t) = \sum_{k=-\infty}^{k=+\infty} s(k\tau) \varphi_k(t).$$

Функции вида

$$\varphi_k(t) = \frac{\sin \Omega(t - k\tau)}{\Omega(t - k\tau)}, \quad k = 0, \pm 1, \dots, \quad \Omega = 2\pi F$$

обладают следующими свойствами:

- 1) интервал, на котором функции $\varphi_k(t)$ ортогональны, равен $(-\infty; +\infty)$,

$$\int_{-\infty}^{\infty} \varphi_k(t) \varphi_n(t) dt = \begin{cases} 0, & \text{если } k \neq n \\ \tau, & \text{если } k = n; \end{cases}$$

- 2) в точке $t = k\tau$ $\varphi_k(t) = 1$, а в точках $t = n\tau$, где n — любое целое, отличное от k , $n \neq k$, функция

$$\varphi_n(t) = 0;$$

- 3) спектр $\tilde{\varphi}_0(\omega)$ равномерен в полосе частот $|\omega| \leq \Omega$ и в указанных пределах равен

$$\tilde{\varphi}_0(\omega) = \frac{1}{2F} = \frac{\pi}{\Omega}.$$

Полученный спектр по виду в точности совпадает с прямоугольным спектром на рис. 3.7, на котором постоянная $C = \frac{1}{2F}$. Так как остальные временные функции $\varphi_k(t)$ отличаются от φ_0 только сдвигом по времени $k\tau$, легко выписать их спектры:

$$\tilde{\varphi}_0(\omega) = \begin{cases} e^{-i\omega k\tau} / 2F = \tau e^{-i\omega k\tau} & \text{при } |\omega| \leq \Omega, \\ 0 & \text{при } |\omega| > \Omega. \end{cases}$$

Модуль функции $\tilde{\varphi}_0(\omega)$ показан на рис. 3.9.

Если взять интервал τ между соседними отсчетами меньшим, чем $1/2F$, $\tau' < 1/2F$, то новое значение τ' интервала будет отвечать большей ширине спектра $(-F'; +F')$, нежели спектр сигнала $s(t)$ (см. рис. 3.9). Этот факт никак не отразится на величине отсчета, совпадающего, как мы видели выше, с коэффициентом ряда Фурье

$$s(k\tau) = \frac{1}{\tau} \int_{-\infty}^{\infty} s(t) \varphi_k(t) dt.$$

Стало быть, сокращение τ допустимо, но технически бесполезно: точность не улучшится, погрешность останется той же — равной нулю, а количество отсчетов увеличится, добавляя работы приемному устройству.

Увеличение τ недопустимо. Спектр $\varphi_k(t)$ при этом сужается (рис. 3.9), становится уже спектра сигнала $s(t)$. Коэффициенты Фурье $s(k\tau)$ при $\tau = \tau \gg \frac{1}{2F''} > \frac{1}{2F}$ уже не являются отсчетами сигнала $s(t)$, а выборочными значениями некоего иного сигнала $s''(t)$, спектр которого ограничен не частотой F , а отличной от нее новой частотой F'' (см. рис. 3.9).

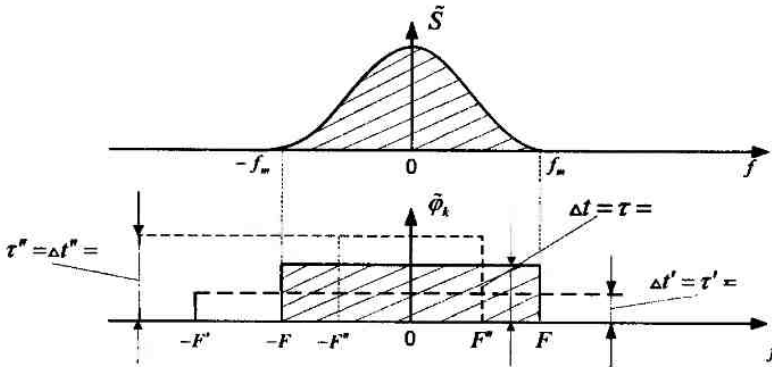


Рис. 3.9. Спектры временных функций $\varphi_k(t)$

3.5.8 Ряд Котельникова при конечном числе отсчетов k

Разложение Котельникова (3.60) записывается для конечного числа отсчетов K , исходя из возможности определения наивысшей частоты спектра Ω , при которой («хвосты») функции $s(t)$, теряющиеся при отсекании частот, превышающих Ω , содержат бы пренебрежимо малую долю энергии по сравнению с энергией исходного сигнала на всем временном интервале

$$E_s = \int_{-\infty}^{\infty} s^2(t) dt = E_s(-\infty; +\infty). \quad (3.75)$$

Определим полную энергию и среднюю мощность сигнала, представленного конечным рядом Котельникова:

$$s_k(t) = \sum_{k=-K}^{k=K} s(k\tau) \varphi_k(t), \quad (3.76)$$

в котором интервал τ , как и в (3.60), равен $1/2F$, а количество отсчетов K подлежит определению.

Отбросим все значения $s(t)$, расположенные вне заданного интервала $[-T; +T]$, т. е. примем, что отсчеты $s(k\tau)$ вне $[-T; +T]$ равны нулю:

$$s(k\tau) = \begin{cases} s(k\tau), & \text{если } -T \leq k\tau \leq +T, \\ 0, & \text{если } k\tau < -T \text{ или } k\tau > T. \end{cases}$$

Новые отсчеты $s(k\tau)$ — это отсчеты ряда (3.76). Тогда максимальное значение K составит

$$K = \lfloor 2F \cdot T \rfloor + 1 = \lceil 2F \cdot T \rceil, \quad (3.77)$$

где скобки $\lfloor \dots \rfloor$ означают взятие целой части числа, находящегося внутри них (операция *entier* в некоторых языках программирования). Иначе говоря, мы зафиксировали тот очевидный факт, что в полуинтервал $[0; T]$ попало не менее $\lfloor 2F \cdot T \rfloor$ и не более $\lceil 2F \cdot T \rceil$ отсчетов. Для гарантии взята верхняя из полученных границ. Получаем усеченный ряд Котельникова

$$s_k(t) = \sum_{k=-\lceil 2FT \rceil}^{+\lceil 2FT \rceil} s(k\tau) \cdot \frac{\sin \Omega(t - k\tau)}{\Omega(t - k\tau)}. \quad (3.78)$$

Величину $K_{\max} = \lceil 2F \cdot T \rceil$ называют базой усеченного сигнала $s_k(t)$ или числом его степеней свободы.

Энергию усеченного сигнала нетрудно выразить, исходя из разложения (3.78) и свойства (3.51) функций $\varphi_k(t)$:

$$E_s(-T; +T) = \sum_{k=-K_{\max}}^{K_{\max}} s^2(k\tau) [\varphi_k(t)]^2 = \tau \cdot \sum_{k=-K_{\max}}^{K_{\max}} s^2(k\tau), \quad (3.79)$$

а мощность выражается следующей очевидной формулой

$$\overline{s^2(t)} = E_s(-T; T)/2T = \frac{1}{4FT} \sum_k s^2(k/2F), \quad (3.80)$$

имеющей смысл среднего квадрата энергий отсчетов на промежутке времени $[-T; T]$.

Осталось теперь сравнить энергию (3.79) с энергией (3.75) исходного сигнала. Если их разность окажется меньше заданного уровня ΔE , например ниже энергетического порога чувствительности фотоприемника, то вместо ряда (3.60) можно использовать в расчетах усеченный ряд (3.78):

$$E_s(-\infty; +\infty) - E_s(-T; T) < \Delta E. \quad (3.81)$$

Практически ΔE составляет, как правило, порядка 1.

Известно много полезных частных случаев теоремы Котельникова, имеющих более простой вид, чем первоначальная. Например, для периодических сигналов $s(t)$ разложение (3.60) удастся записать с помощью конечной суммы, несмотря на то что область определения сигнала $s(t)$ остается бесконечной. Можно показать, что если $[0; T]$ — период s_t и сигнал s_t не содержит частот выше $F = K/T$, то для $t \in (-\infty; \infty)$ справедливо представление

$$s(t) = \sum_{k=0}^{2K} s \left(\frac{kT}{2k+1} \right) \varphi \left(t - \frac{kT}{2k+1} \right),$$

при этом

$$\varphi(t) = \frac{\sin[(2k+1)(\pi t/T)]}{(2k+1) \sin(\pi t/T)}.$$

Интересен случай распространения теоремы на эквивалентные сигналы. Эквивалентными называют сигналы $u(t)$ и $v(t)$ на $t \in (-\infty; +\infty)$, имеющие совпадающие ряды отсчетов:

$$u(k\tau) = v(k\tau) \quad \text{для } k = 0, \pm 1, \pm 2, \dots$$

Пусть $\tilde{u}(f), \tilde{v}(f)$ — спектры сигналов $u(t)$ и $v(t)$, зависящие от частоты $f = \omega/2\pi$. Тогда

$$\sum_{k=-\infty}^{\infty} \tilde{u}\left(f - \frac{k}{\tau}\right) = \sum_{k=-\infty}^{\infty} \tilde{v}\left(f - \frac{k}{\tau}\right), \quad \text{где } f \in (-\infty; +\infty).$$

Помимо того, если $u(0) = 1$ и $u(k\tau) = 0$ для всех $k \neq 0$ (сигнал с «выколотыми» значениями), то для всех $f \in (-\infty; +\infty)$

$$\sum_{k=-\infty}^{\infty} \tilde{u}\left(f - \frac{k}{\tau}\right) = \tau.$$

Иначе говоря, любые «выколотые» сигналы, регулярно обращающиеся в нуль через интервал τ и ограниченные по спектру частотой $F = \frac{1}{2\tau}$, имеют относительно низкую суммарную спектральную мощность (порядка той, которой обладают временные функции $\varphi_k(t)$).

3.5.9 Теорема отсчетов в частотной области (теорема 2)

Иногда спектр сигнала $\tilde{s}(\omega)$ нужно представить дискретными отсчетами, например когда приемопередающая аппаратура строится с использованием цифровых фильтров и компьютерных микросхем памяти. Для спектра $\tilde{s}(\omega)$ можно составить ряд, аналогичный (3.60) или (3.78).

Для спектра $\tilde{s}(f)$ может быть получена теорема 2.

Теорема 2. Для сигнала $s(t)$, ограниченного на $|t| \leq T$ и принимающего вне этого интервала значения $s(t) = 0$, непрерывный спектр $\tilde{s}(f)$ можно представить рядом

$$\begin{aligned} \tilde{s}(f) &= \sum_{l=-\infty}^{\infty} \tilde{s}\left(\frac{2\pi l}{2T}\right) \cdot \frac{\sin 2\pi T\left(f - \frac{l}{2T}\right)}{2\pi T\left(f - \frac{l}{2T}\right)} = \\ &= \sum_{l=-\infty}^{\infty} \tilde{s}(l2\pi\Delta f) \cdot \frac{\sin 2\pi T(f - l\Delta f)}{2\pi T(f - l\Delta f)} = \\ &= \sum_{l=-\infty}^{\infty} \tilde{s}_s(l\Delta\omega) \cdot \frac{\sin [T(\omega - l\Delta\omega)]}{T(\omega - l\Delta\omega)}, \end{aligned} \quad (3.82)$$

где Δf — шаг отсчета частоты.

Теорема 2 называется «дуальной» теореме 1.

Имеет смысл записать дуальный ряд (3.82) в конечном виде как функцию круговой частоты ω , которая дискретизируется с интервалом $\Delta\omega = \pi/T$. При ширине спектра 2Ω , охватывающей область частот $-\Omega < \omega < +\Omega$, количество отсчетов в конечном ряду равно $2\Omega/\Delta\omega = 2 \cdot 2\pi F/(\pi/T) = 4FT$. Спектральные

отсчеты $\tilde{s}(l\Delta\omega)$, будучи комплексными числами, удовлетворяют условию

$$\tilde{s}^* \left(-l \frac{\pi}{T} \right) = \tilde{s} \left(l \frac{\pi}{T} \right), \quad (3.83)$$

где $*$ — знак комплексной сопряженности, так что в (3.82) можно пользоваться только положительными частотами.

Имеем

$$\tilde{s}(\omega) = \sum_{l=-4FT}^{4FT} \tilde{s}(l\Delta\omega) \cdot \frac{\sin[T(\omega - l\Delta\omega)]}{T(\omega - l\Delta\omega)}. \quad (3.84)$$

Спектральные отсчеты в конечном спектре на интервале $[-\Omega; +\Omega]$ показаны на рис. 3.10.

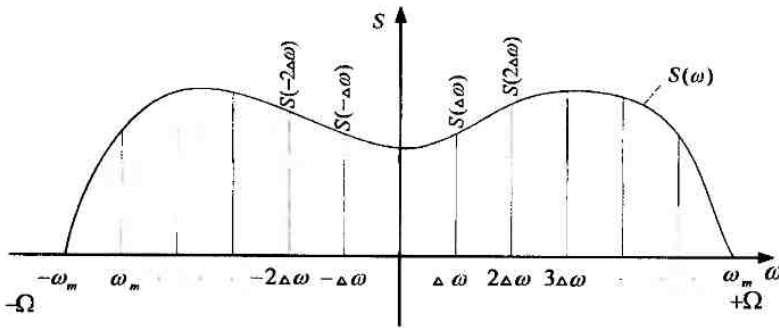


Рис. 3.10. Спектральные отсчеты

Из рис. 3.10 хорошо видно, что спектр сигнала полностью характеризуется комплексными отсчетами, взятыми только в области положительных частот, а число степеней свободы составляет $4FT$. При этом вещественные части отсчетов четны относительно $\omega = 0$, а мнимые составляющие — нечетны в силу условия (3.83).

Существует также любопытное дуальное Фурье-разложение спектра

$$\tilde{s}(f) = \frac{1}{2F} \sum_{l=-\infty}^{\infty} \tilde{s}\left(\frac{l}{2F}\right) e^{-\frac{i\pi l f}{F}}, \quad (3.85)$$

где $|f| \leq F$ и

$$\tilde{s}\left(\frac{l}{2F}\right) = \int_{-F}^{+F} \tilde{s}(f) e^{-\frac{i\pi l f}{F}} df. \quad (3.86)$$

3.5.10 Обобщение теоремы Котельникова на случай произвольной полосы частот ширины F

Пусть полоса

$$[\Omega_0 - \pi F; \Omega_0 + \pi F] \quad (3.87)$$

имеет своим центром частоту Ω_0 и полную ширину $2\pi F$.

Сигнал $S(t)$, частотный спектр $\tilde{S}(w)$ которого отличен от нуля только в пределах ограниченной полосы частот, как нетрудно показать, может быть представлен в виде

$$S(t) = SS(t) \sin \Omega_0 t + SC(t) \cos \Omega_0 t, \quad (3.88)$$

где $SS(t)$ и $SC(t)$ — функции, имеющие отличные от нуля спектры $\tilde{SS}(w)$ и $\tilde{SC}(w)$ только для диапазона

$$|w| < \pi F. \quad (3.89)$$

Согласно известным свойствам Фурье-преобразований спектры $\tilde{S}(w)$, $\tilde{SS}(w)$ и $\tilde{SC}(w)$ связаны соотношением

$$\tilde{S}(w) = \frac{1}{2}[\tilde{SC}(w - \Omega_0) + i\tilde{SS}(w - \Omega_0)] + \frac{1}{2}[\tilde{SC}(w + \Omega_0) + i\tilde{SS}(w + \Omega_0)]. \quad (3.90)$$

Поскольку $\tilde{SC}$ и $\tilde{SS}$ отличны от нуля только при $|w| \leq \pi F$, то $\tilde{S}(w)$ может отличаться от нуля только при условии

$$w \in [\Omega_0 - \pi F; \Omega_0 + \pi F]. \quad (3.91)$$

Итак, $SS(t)$ и $SC(t)$ — функции с ограниченным спектром, которые можно разложить в ряды

$$SS(t) = \sum_{k=-\infty}^{\infty} SS(k\tau) \frac{\sin \pi F(t - k\tau)}{\pi F(t - k\tau)}, \quad (3.92)$$

$$SC(t) = \sum_{k=-\infty}^{\infty} SC(k\tau) \frac{\sin \pi F(t - k\tau)}{\pi F(t - k\tau)}, \quad (3.93)$$

где $SS(k\tau)$, $SC(k\tau)$ — амплитудные отсчеты.

По терминологии, принятой в технике связи, два члена в правой части равенства (3.88) получаются при модуляции сигнала с подавлением несущей, когда опорные гармонические несущие $\sin \Omega_0 t$ и $\cos \Omega_0 t$ модулируются сигналами $SS(t)$ и $SC(t)$.

Обобщение теоремы Котельникова на полосу частот $\Omega_0 \pm \pi F$ выглядит следующим образом.

Теорема 3. Сигнал $S(t)$, строго ограниченный по полосе частот $\Omega_0 \pm \pi F$, можно восстановить по формулам (3.88), (3.91), (3.92), при этом амплитудные отсчеты $SS(k\tau)$, $SC(k\tau)$, $k = 0, \pm 1, \pm 2, \dots$ осуществляются через равноотстоящие промежутки времени с шагом

$$\tau = \frac{1}{2F}.$$

Иначе говоря, по отсчетам $SS(k\tau)$ и $SC(k\tau)$ восстанавливается не сам сигнал, а его огибающая.

Для практического применения (3.88), впрочем, как и для всех предыдущих разложений, отсчеты нормируются таким образом, чтобы средняя мощность сигнала S была бы равна сумме квадратов отсчетов:

$$\overline{S^2} = \frac{1}{2\tau} \int_{-T}^{+T} S^2(t) dt = \sum_{k=-2FT}^{k=2FT} S_k^2. \quad (3.94)$$

Для этого коэффициенты $S(k\tau)$ в рядах Котельникова следует разделить на $\sqrt{2FT}$, так как энергия каждого члена ряда равна $S_k^2/2F$.

Аналогично, коэффициенты $SS(k\tau)$ и $SC(k\tau)$ в разложении (3.88) следует разделить на $\sqrt{FT}$.

Нормированные отсчеты $S_1, S_2, \dots, S_{2TF}$ можно рассматривать как декартовы координаты точки, представляющей сигнал $S(t)$. Средняя мощность такого сигнала совпадает с квадратом расстояния соответствующей точки от начала координат $2TF$ -мерного пространства. Скалярное произведение двух таких векторов равно взаимной корреляции сигналов.

Весьма полезна теорема 3 в задачах представления узкополосных сигналов, для которых

$$2\pi F \ll \Omega_0 \quad (3.95)$$

(критерий узкополосности).

Узкополосный сигнал можно представить в виде

$$S(t) = A(t) \cos[\Omega_0 t + \theta(t)], \quad (3.96)$$

при этом его спектр $\tilde{S}(w)$ показан на рис. 3.10, а $A(t)$ — огибающая, которую можно разложить в ряд. Если интервал τ между отсчетами огибающей $A(t)$ не превышает $1/F$, то ряд Котельникова для нее имеет вид

$$A(t) = \sum_{k=-\infty}^{\infty} A(k\tau) \cdot \frac{\sin 2\pi F(t - k\tau)}{2\pi F(t - k\tau)}, \quad (3.97)$$

откуда для $S(t)$ имеет разложение

$$S(t) = \sum_{k=-\infty}^{\infty} A(k\tau) \cdot \frac{\sin 2\pi F(t - k\tau)}{2\pi F(t - k\tau)} \cdot \cos[\Omega_0 t + \theta(t)]. \quad (3.98)$$

С помощью полосового разложения (3.97) можно оценивать количество информации, содержащееся в модулированных сигналах, сопоставлять по информации сигналы с различными видами модуляции.

3.5.11 Теорема отсчетов для АМ и ЧМ сигналов

Амплитудная модуляция — это изменение во времени $A(t)$, спектр которой $\tilde{A}(w)$ содержит максимальную частоту $\Omega = 2\pi F$ (см. (3.91)). Ширина спектра $S(t)$ равна $\Delta f_{\text{ам}} = 2F$, причем в пределах этой полосы спектр $\tilde{S}(w)$ симметричен относительно Ω_0 . Интервал τ между отсчетами $A(k\tau)$ должен быть не более $1/\Delta f_{\text{ам}}$, т. е. таким же, как и при дискретизации исходного сигнала (модулирующей функции).

Фазу $\Omega_0 t + \theta(t)$ при АМ передавать по каналу нет необходимости. Отсюда получаем еще одну полезную теорему.

АМ сигнал вполне определяется отсчетами своей амплитуды, взятыми через $1/\Delta f_{\text{ам}}$, где $\Delta f_{\text{ам}}/2$ — верхняя частота в спектре модулирующей функции

(т. е. передаваемого сообщения). Иначе говоря, число степеней свободы АМ сигнала такое же, как и число степеней свободы модулирующей функции.

Рассмотрим теперь ЧМ сигнал

$$S(t) = A_0 \cos[\Omega_0 t + \theta(t)], \quad (3.99)$$

в котором мгновенная частота $w(t) = \Omega_0 + d\theta(t)/dt$ модулирована тем же сообщением, что и выше. При этом примем, что максимальная девиация частоты $f_{\text{дев}} \gg F$. Тогда ширину полосы частот $\Delta f_{\text{ЧМ}}$ модулированного колебания можно приравнять $2f_{\text{дев}}$:

$$\Delta f_{\text{ЧМ}} = 2f_{\text{дев}}. \quad (3.100)$$

Случай (3.99) сигнала (3.98) иногда именуют широкополосной частотной модуляцией. Передавать A_0 при ЧМ нет необходимости. Достаточно передать отсчеты фазы $\theta(k\tau)$, $k = 0, \pm 1, \dots$ в моменты, отстоящие друг от друга на $\tau \leq 1/2f_{\text{дев}}$.

Получается, что при одной и той же длительности сигнала $2T$ число отсчетов фазы при ЧМ равно

$$K_{\Phi} \Delta f_{\text{ЧМ}} \cdot 2T = 4f_{\text{дев}} \cdot T, \quad (3.101)$$

а количество отсчетов огибающей при АМ равно

$$K_a = \Delta f_{\text{АМ}} \cdot 2T = 4F \cdot T. \quad (3.102)$$

Отсюда следует, что при одном и том же передаваемом сигнале (информационном сообщении) ЧМ сигнал обладает в $\mu = f_{\text{дев}}/F$ раз большим числом степеней свободы, чем АМ сигнал. На приемном конце канала связи после частотного детектирования можно сформулировать напряжение, имеющее спектр исходного сигнала. Итак, при одной и той же ширине спектра $2F$ информационное содержание ЧМ сигнала выше.

ЧМ сигнал вполне определяется отсчетами своей фазы $\theta(\tau)$, взятыми через интервал τ ,

$$\tau \ll 1/2f_{\text{дев}}, \quad (3.103)$$

где $f_{\text{дев}}$ — частотная девиация, равная половине ширины полосы частот модулированного сигнала. При смешанной АМ-ЧМ модуляции в каждой временной точке нужно брать два отсчета: амплитуды A и фазы θ .

Изучение ВИМ (временнo-импульсно модулированных) сигналов можно свести к изучению ЧМ сигналов. ВИМ сигналы так же восстанавливаются с помощью фазы $\theta(\tau)$, причем интервал τ можно варьировать в еще больших пределах, чем при ЧМ, повышая информационную емкость сигналов.

Приемопередающий тракт, базирующийся на ОАК, работает в супервысоко-частотном импульсном режиме, для которого режим ВИМ наиболее естественен и технически прост.

3.5.12 Неравномерные отсчеты

Пользоваться равномерными отсчетами амплитуды, частоты, фазы сигналов не всегда удобно. Так, ввиду большой информационной емкости оптического диапазона целесообразно организовать многоканальный режим работы канала связи

с временным делением каналов, когда каждому абоненту периодически выделяется относительно небольшой промежуток времени. Такой режим требуется и в системах телеметрии, адресной передачи данных по компьютерным сетям. Если продолжать пользоваться дискретными отсчетами и ВИМ, то необходимо обобщить теорему Котельникова на случай отсчетов, которые не обязательно следуют равномерно друг за другом.

Рассмотрим случай, в котором отсчеты распределены на оси времени группами по N точек в каждой. Группы отсчетов периодически повторяются, расположение отсчетов внутри группы произвольно. Отобразим описанное групповое расположение на множество равномерных отсчетов, взяв для простоты случай трехточечной группы, $N = 3$ (рис. 3.11). Обозначим t_{kl} — моменты отсчетов:

$$t_{kl} = Nk\tau + \tau_l, \quad (3.104)$$

где $k = 0, \pm 1, \pm 2, \dots$ (как и выше);

$$l = 0, 1, 2, \dots, N - 1;$$

$$\tau_0 = 0; \tau_l < N\tau;$$

$$\tau = \frac{1}{2F};$$

при каждом фиксированном l (т.е. при выборе определенной точки τ_l внутри группы) отсчеты

$$\dots, t_{-2l}, t_{-1l}, t_{0l}, t_{1l}, t_{2l}, \dots$$

образуют последовательность равноотстоящих значений. Для рис. 3.11 имеем

$$N = 3; t_{k0} = 3k\tau,$$

$$t_{kl} = 3k\tau + \tau_1,$$

$$t_{k2} = 3k\tau + \tau_2,$$

при этом последовательность t_{k0}, t_{k1}, t_{k2} имеет период $T_N = 3\tau$.

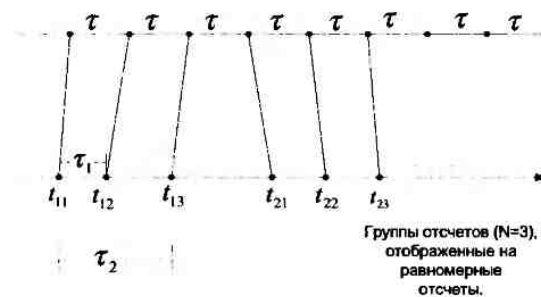


Рис. 3.11. Группы отсчетов ($N = 3$), отображенные на равномерные отсчеты

3.5.13 Теорема для сгруппированных отсчетов

Дальнейшие действия по построению ряда Котельникова базируются на соответствии равномерных и сгруппированных отсчетов (рис 3.11). За время $N\tau$ на верхней и нижней осях времени берется одинаковое число N отсчетов. Количество информации при первом (равномерном) и втором (неравномерном) способах

формирования последовательностей отсчетов, очевидно, не должны сильно отличаться. Поэтому можно ожидать, что для функций $S(t)$ с конечным спектром $\pm 2\pi F$ и $\tau \leq \frac{1}{2F}$ справедлива интерполяционная формула.

Теорема 4.

$$S(t) = \sum_{k=-\infty}^{\infty} \sum_{l=0}^{N-1} S(t_{kl}) \cdot \varphi_N(t; \tau), \quad (3.105)$$

где

$$\varphi_N(t; \tau) = \frac{\prod_{m=0}^{N-1} \sin \left[\frac{\pi}{N\tau} (t - t_{km}) \right]}{\frac{\pi}{N\tau} (t - t_{kl}) \cdot \prod_{m=0, m \neq l}^{N-1} \sin \frac{\pi}{N\tau} (\tau_k - \tau_m)}, \quad (3.106)$$

при этом, подобно (3.60), функция $\varphi_N(t; \tau) = 1$ при $t = t_{kl}$, а при $t = t_{mr}$, где $m \neq k$ и $r \neq l$, обращается в нуль, так что $S(t_{kl}) =$ правой части (3.105). Ряд (3.105) сходится равномерно, т. е. для любого t из ограниченной области.

Формулы (3.105), (3.106) составляют запись теоремы Котельникова для неравномерных отсчетов, распределенных по периодически повторяющимся группам из N отсчетов в каждой, при выполнении условий (3.104).

3.5.14 Теорема для неравномерных достаточно регулярных отсчетов

Существуют обобщения теоремы (3.105) на неравномерные отсчеты, на которые наложены достаточно общие необременительные условия. Одно из этих обобщений формулируется так.

Теорема 5. Пусть последовательность $(t_k)_{-\infty}^{\infty}$ моментов отсчетов удовлетворяет условиям:

а) для всех $k = 0, \pm 1, \pm 2, \dots$

$$t_{k+1} - t_k > c > 0; \quad (3.107)$$

б) среднее значение интервала между отсчетами равно τ :

$$\lim_{k \rightarrow \infty} [(t_k - t_{-k})/2k] = \tau \quad (3.108)$$

(условие равномерной асимптотической плотности моментов отсчетов).

Тогда для сигнала $S(t)$ с ограниченным спектром, $|w| \leq 2\pi F = \frac{\pi}{\tau}$, справедливо разложение

$$S(t) = \sum_{k=-\infty}^{\infty} \frac{S(t_k) \cdot \varphi(t)}{\varphi'(t_k)(t - t_k)}, \quad (3.109)$$

в котором

$$\varphi(t) = \prod_{l=-\infty}^{k-1} \sin \frac{\pi}{k\tau} (t - t_l). \quad (3.110)$$

В теореме 5 важнее всех условие равномерной плотности отсчетов (3.108). Если отсчеты разбросаны по оси времени достаточно регулярно, не образуют скоплений и сгущений, то обобщенное разложение (3.110) Котельникова остается полезным способом дискретизации аналоговых сигналов в канале связи.

Математиками, использующими продвинутое методы теории функций комплексного переменного, получено несколько десятков важных частных случаев последних теорем (например, для фиксированных численностей групп $N = 2, 3, \dots, 5$). В разложениях последних лет участвуют не только функции времени и спектры, но и их производные. Для последних приходится строить особую процедуру формирования.

За истекшие 15–20 лет более тщательно изучены вопросы оценивания погрешности представления и восстановления аналоговых сигналов, определенных на конечных подмножествах временной оси t . Основательно проработаны условия сходимости бесконечных рядов из дискретных отсчетов, определен класс фильтров, восстанавливающих сигнал $S(t)$ на приемном конце канала с требуемой точностью в условиях стационарных случайных помех.

Для примера можно привести обобщенное разложение Котельникова сигнала $S(t)$, ограниченного на $(-\infty; +\infty)$ по мощности, для частного случая $N = 2$ (периодически повторяется группа из двух неравномерных отсчетов).

Теорема 6. Если спектр $\tilde{S}(f)$ сигнала $S(t)$ ограничен полосой $|f| \leq F$ и характерная средняя длительность интервала между отсчетами составляет $\tau \leq \frac{1}{2F}$, то при достаточно малых $\tau \rightarrow 0$ получается предельное равенство

$$S(t) = \sum_{k=-\infty}^{\infty} [S(2k\tau) + (T - 2K\tau) \cdot S'(2k\tau)] \cdot \left[\frac{\sin \frac{\pi}{2\tau}(t - 2k\tau)}{\frac{\pi}{2\tau}(t - 2k\tau)} \right]^2. \quad (3.111)$$

Для более общих случаев с произвольным количеством неравномерных отсчетов N найдены полиномиальные передаточные функции полосовых фильтров, которые можно использовать для восстановления аналоговых сигналов при комплексе заданных технических условий.

3.6 Теоремы Найквиста и Шеннона

3.6.1 Формула Найквиста

Первоначальная формула Найквиста (Nyquist Н.) описывает частотное распределение случайных флуктуаций тока или напряжения в замкнутой электрической цепи. В 1928 году Х. Найквист показал [181], что указанные флуктуации можно представить как результат действия случайной э.д.с. E , среднее временное значение $\overline{E}$ которой равно нулю:

$$\overline{E} = \frac{1}{2\tau} \int_{-\tau}^{\tau} E(t) w(E, t) dE dt = 0 \quad (3.112)$$

для любого заданного периода времени τ , в том числе $\tau \rightarrow \infty$, где t — время, $w(E, t)$ — плотность распределения вероятностей э.д.с. $E(t)$, а среднее квадратичное ее частотной компоненты спектра $E(\omega)$ в частотном интервале от ω до $\omega + d\omega$ равно

$$\sigma^2[E(\omega)] = e^2(\omega) = k_B T R / \pi, \quad (3.113)$$

где k_B — постоянная Больцмана, $R = \operatorname{Re}(Z)$, $Z = Z(\omega)$ — комплексное сопротивление цепи, T — абсолютная температура в градусах Кельвина. Период времени в реальных задачах берется достаточно большим, чтобы все переходные процессы в цепи уже закончились. Поэтому без ограничения общности в формуле (3.112) пределы интегрирования можно считать равными $\pm\infty$, а саму формулу понимать в предельном смысле:

$$\bar{E} = \lim_{\tau \rightarrow \infty} \left[\frac{1}{2\tau} \int_{-\tau}^{\tau} E(t) \omega(E, t) dE dt \right]. \quad (3.114)$$

Флуктуации можно представить как результат действия генератора тока с бесконечно большим внутренним сопротивлением. Генератор создает флуктуирующий ток с нулевым средним значением и средним квадратичным

$$\sigma^2[I(\omega)] = \tau^2(\omega) = k_B T R / (\pi \cdot |z|^2). \quad (3.115)$$

В современной литературе по радиофизике формулами Найквиста называют равенства (3.113) и/или (3.115). В квантовой механике получена более общая формула, также носящая имя Найквиста, частным случаем которой являются (3.113) и (3.115):

$$e^2(\omega) = \frac{1}{2} \hbar \omega / (\pi R \operatorname{cth}(\hbar \omega / 2k_B T)). \quad (3.116)$$

Следствия (3.113) и (3.115) вытекают из (3.116) при

$$k_B T \gg \hbar \omega, \quad (3.117)$$

т. е. при достаточно высоких абсолютных температурах

$$T > T_0 = 10 \cdot \frac{\hbar \omega}{k_B} = \frac{10 \cdot (1,0545919 \pm 0,0000080) \cdot 10^{-34} [\text{Дж} \cdot \text{с}]}{(1,380622 \pm 0,000059) \cdot 10^{-23} [\text{Дж/град}]} \cdot \omega = \\ = (0,763853 \pm 0,000034) \cdot 10^{-10} \omega [\text{град}].$$

Для $T \approx 300\text{K}$ величина $\hbar \omega / 2k_B T$ составляет приблизительно 10^3 ; при этом с точностью не хуже 10^{-12} $\operatorname{cth}(\hbar \omega / 2k_B T)$ равен 1, так что формула (3.116) приобретает простой вид

$$e^2(\omega) = \frac{1}{2} \hbar \omega / \pi \operatorname{Re}(Z), \quad (3.118)$$

где $\omega \in (0, 6 \dots 3,6) \cdot 10^{15} \frac{\text{рад}}{\text{с}}$, а Z — волновое комплексное сопротивление среды, для которой построена эквивалентная схема.

Формула (3.116) получена для пассивной цепи, построенной из емкостей, индуктивностей и сопротивлений. Цепь может включать и другие пассивные элементы, например электромеханические, электроакустические и другие интегральные элементы при условии, что все они находятся при одинаковой температуре T . Каждый элемент заменяется его эквивалентной схемой. При этом емкости и прочие элементы являются достаточно большими, чтобы их усредненные по объему характеристики могли выполнять функции макропараметров усредненных уравнений Максвелла, и в то же время достаточно малыми, чтобы для них, как для кристаллических структур, оставались справедливыми формулы квантовой механики.

Если пользователь изучает влияние помех на несущих частотах порядка $F = 10^{14...15}$ Гц, то он обязан пользоваться формулой (3.117). Однако в аппаратуре систем связи с применением ОАК, в частности во входных и выходных цепях микросхем приемопередатчиков, используются частоты $F \simeq 10^9...10^{10}$ Гц, для которых выполнено условие (3.117), и можно применять классические формулы Найквиста (3.113) и (3.115).

Изучим классические формулы Найквиста более подробно в связи с изложенной выше теоремой Котельникова, ограничившись рассмотрением идеальных сигналов в виде прямоугольных импульсов ВИМ фиксированной амплитуды A , длительности Δt , следующих с периодом ΔT . Отрезок времени ΔT , в течение которого возможно появление импульса, именуется, как и выше, позицией с тактом N примыкающих позиций, $8 \leq N \leq 32$, и образует кадр длительности $T_k = N\Delta T$ с коэффициентом заполнения $c = \Delta\tau/\Delta T$.

Классическая формула (3.113) исходит из законов случайного движения классического электрона как заряженной частицы в электрической цепи. Это движение сходно с броуновским движением молекул идеального газа в среде с определенным коэффициентом вязкости.

Пусть электрон (ион) массы m под действием случайной силы $E(t)$ движется в среде с коэффициентом вязкого трения α со скоростью $\dot{x}$

$$m\ddot{x} + \alpha\dot{x} = E. \quad (3.119)$$

Обозначим τ — промежуток времени между столкновениями электрона (иона) с отталкивающими центрами среды. Пусть промежуток времени τ достаточно мал, так что изменение скорости $\Delta\dot{x} = \dot{x}(t_1) - \dot{x}(t_0)$, $t_1 - t_0 = \tau$, также мало по отношению к $|\dot{x}|$. Отсюда, заменяя $\ddot{x}(t) = \Delta\dot{x} / \tau$, будем иметь:

$$m\dot{x}(t_1) = \dot{x}(t_0)[m - \alpha\tau] = X, \quad (3.120)$$

где $X = \int_0^\tau E(t)dt$. Так как $E(t)$ — случайная величина, то

$$\overline{X} = \int_{-\infty}^{\infty} E(t)dt,$$

но при этом интенсивность $\overline{X^2}$ не равна нулю. Возводя левую и правую части (3.120) в квадрат, усредняя и пренебрегая малыми членами,

$$\alpha^2\tau^2/m^2 \ll \frac{2\alpha\tau}{m}, \quad (3.121)$$

придем, положив $X = 0$, к равенству

$$\overline{\dot{x}^2(t_1)} = \overline{\dot{x}^2(t_0)} \left[1 - \frac{2\alpha\tau}{m} \right] + \frac{\overline{X^2}}{m^2}. \quad (3.122)$$

Поскольку система «электрон—среда» находится в равновесном состоянии и согласно молекулярно-кинетической теории имеет установившуюся температуру T , то кинетическая энергия

$$\frac{1}{2}m\dot{x}^2(t) = \frac{1}{2}k_B T. \quad (3.123)$$

Так как в равновесном состоянии $\overline{\dot{x}^2(t_1)} = \overline{\dot{x}^2(t_0)} = k_B T/m$, из равенства (3.122) получим любопытную формулу

$$\overline{X^2} = 2\alpha\tau k_B T. \quad (3.124)$$

Умножая левую и правую части (3.122) на $\frac{1}{2}m$, приходим к соотношению между энергиями

$$\frac{1}{2}m\overline{\dot{x}^2(t_1)} = \frac{1}{2}m\overline{\dot{x}^2(t_0)} \left[1 - \frac{2\alpha\tau}{m} \right] + \frac{\overline{X^2}}{m^2}, \quad (3.125)$$

в котором левая часть — сумма кинетических энергий движущейся частицы и случайного электрического поля, а величина

$$\frac{1}{2}m\overline{\dot{x}^2(t_0)} \cdot \frac{2\alpha\tau}{m}$$

представляет собой потери кинетической энергии частицы при ее движении через кристаллическую решетку. Тогда $\overline{X^2}/2m$ — энергия, которая доставляется движущейся частице полем и случайными столкновениями с колеблющимися центрами. Получается, что работа случайного поля производится за счет теплового движения и вновь превращается в теплоту с помощью сил вязкого трения, зависящих от скорости движения частицы.

С помощью формулы (3.122) можно получить выражение для среднего квадрата пространственного смещения частицы $\overline{x^2(t)}$

$$\overline{x^2(t)} = \overline{X^2}/\alpha^2,$$

откуда

$$\overline{x^2(t)} = \frac{2k_B T \cdot t}{\alpha}, \quad (3.126)$$

где t — время, $t \ll \tau$. Формула (3.126) неоднократно проверялась экспериментально по движению коллоидных частиц в суспензии, заряженных капель, ионов и электронов в металлах и полупроводниках. Правильность (3.126) косвенно подтверждает и справедливость (3.124).

Если теперь немного изменить условия и рассмотреть цепь из сопротивления R и индуктивности L , то для тока I получится уравнение

$$L \frac{dI}{dt} + RI = E(t), \quad (3.127)$$

где E — случайная э.д.с. в цепи (рис. 3.12).

Если цепь находится при постоянной температуре T , то

$$\frac{1}{2}L\overline{I^2(t)} = \frac{1}{2}k_B T. \quad (3.128)$$

Уравнение (3.127) — копия уравнения (3.119), условие (3.128) аналогично (3.123). Подставляя R вместо α в (3.124), получим формулу для среднего квадрата случайной э.д.с.:

$$\overline{X^2} = 2R\tau k_B T.$$

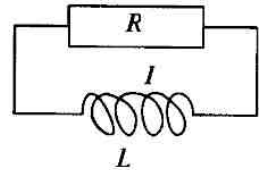


Рис. 3.12. Цепь тока (3.129)

Если случайную э.д.с. представить как множество импульсов случайной амплитуды и малой длительности $\Delta\tau$ (рис. 3.13), то находим

$$X = \int_{t=0}^{t=N\Delta\tau} E(t)dt = \Delta\tau(E_1 + E_2 + \dots + E_N)$$

и

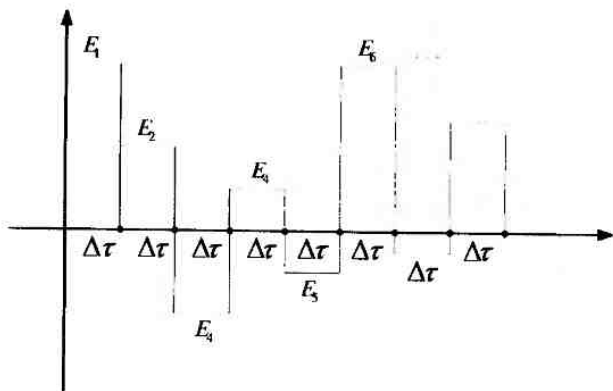
$$X^2 = (\Delta\tau)^2(E_1^2 + E_2^2 + \dots + E_N^2 + 2E_1E_2 + 2E_1E_3 + \dots).$$

После усреднения последней суммы получаем:

$$\overline{X^2} = (\Delta\tau)^2(\overline{E_1^2} + \overline{E_2^2} + \dots),$$

так как средние значения смешанных произведений равны нулю в силу статистической независимости импульсов.

Рис. 3.13. Импульсы случайного поля



В итоге

$$\overline{X^2} = (\Delta\tau)^2 \overline{E^2} = \tau \overline{E^2 \Delta\tau} \quad (3.130)$$

средний квадрат импульса силы (или средний квадрат э.д.с.) возрастает пропорционально времени.

Из формулы $\overline{X^2} = \tau \cdot \Delta\tau \overline{E^2} = 2R\tau kT$, которая вытекает из (3.129) и (3.130), имеем

$$\overline{E^2} \Delta\tau = 2RkT. \quad (3.131)$$

Осталось сделать несколько заключительных шагов.

Для одиночного импульса рис. 3.13 амплитуды E и длины $\Delta\tau$ имеем спектральное представление

$$s(t) = \int_{-\infty}^{\infty} \tilde{s}(f) \cos 2\pi f t df = \int_0^{\infty} 2\tilde{s}(f) \cos 2\pi f t df, \quad (3.132)$$

в котором, как показано ранее, спектральная плотность имеет вид

$$\tilde{s}(f) = E\Delta\tau \frac{\sin \pi f \Delta\tau}{\pi f \Delta\tau}. \quad (3.133)$$

Ширина полосы Δf частот для прямоугольного импульса длительности $\Delta\tau$ составляет

$$\Delta f = 2f_{\max} = \frac{1}{\Delta\tau}, \quad -f_{\max} \leq f \leq +f_{\max}, \quad f_{\max} = \frac{1}{2\Delta\tau}.$$

В самом грубом приближении можно считать, что спектр прямоугольного импульса амплитуды E и длительности $\Delta\tau$ постоянен на интервале $(-f_{\max}; f_{\max})$,

$$\tilde{s}(f) = E\Delta\tau.$$

Если вводить отрицательные частоты по каким-либо причинам нежелательно, можно положить

$$0 \leq f \leq f_{\max} = \frac{1}{2\Delta\tau},$$

удвоив спектральную плотность

$$\tilde{s}(f) = 2\tilde{s}(f) = 2E\Delta\tau. \quad (3.134)$$

Средний квадрат э.д.с. равен, таким образом,

$$\overline{E^2} = 2Rk_B T / \Delta\tau \quad (3.135)$$

и занимает полосу частот

$$0 \leq f \leq f_{\max},$$

поэтому

$$\overline{E^2(f)} = 4Rk_B T f_{\max}. \quad (3.136)$$

В последней формуле можно сделать пропорциональный переход к меньшему интервалу частот:

$$\overline{e^2} = 4Rk_B T \Delta f. \quad (3.137)$$

Это и есть искомая формула Найквиста для замкнутого контура с активным сопротивлением R .

Нетрудно видеть, что помехи возникают только в сопротивлении R , т.е. в элементе, рассеивающем энергию, в котором наблюдается связь между током и тепловым движением. Индуктивности и емкости не дают помех! Этот замечательный и глубокий вывод распространяется на достаточно широкий класс физических систем и сред, в том числе и на молекулярные стохастические системы, одной из которых и является ОАК.

Главное достоинство формулы Найквиста — ее простота и универсальность. Она позволяет давать оценки интенсивностей помех, точные не только по порядку величины, но и фактически.

3.6.2 Обобщения формулы Найквиста

Пусть X — импульс случайной силы за время t , а R — коэффициент затухания электромагнитных волн в атмосфере. Сила эта равна энергии флуктуирующей

электромагнитной (световой) рассеиваемой волны W , деленной на диаметр d характерной области рассеяния. Чтобы получить импульс X , нужно частное W/d умножить на время t (т. е. э.д.с. умножается на t). Среднее квадратичное отклонение импульса (энергия помех) равно коэффициенту затухания световой волны, умноженному на $k_B T \cdot t$, где $k_B T$ — энергия молекулы воздуха, а t — время, в течение которого существует световая волна в области диаметром d .

Отсюда получаем, что средний квадрат сил $\overline{X^2}$, возникающих вследствие рассеяния на молекулах воздуха электромагнитных волн достаточно низких частот (не более ТГц), равен

$$\overline{X^2} = 4Rk_B T \Delta f, \quad (3.138)$$

где Δf — ширина спектра импульса. При этом $k_B T$ — средняя энергия элементарных осцилляторов электромагнитного излучения.

Если условие $\hbar\omega \ll kT$ нарушено, среднюю энергию осцилляторов следует взять равной не $k_B T$, а

$$\overline{e(\omega)} = \frac{\hbar\omega}{e^{\frac{\hbar\omega}{k_B T}} - 1} \quad (3.139)$$

Во всех формулах следует kT заменить на $\overline{e(\omega)}$.

В 1930–1960 годах 20 века были предприняты большие усилия по проверке приведенных выше формул, в том числе по применению к ряду неэлектромагнитных задач при ограничениях самого различного вида.

3.6.3 Формула Шеннона–Таллера–Хартли

Формула Шеннона определяет пропускную способность канала связи (т. е. некой физической среды) с учетом помех.

Для обсуждения этой формулы напомним сначала, что низкочастотные помехи характеризуются средней энергией

$$\overline{W} = f_{\max} \cdot \tau \cdot k_B \cdot T, \quad (3.140)$$

где τ — длительность сигнала и $f_{\max}$ — максимальная частота в его спектре, и мощностью

$$P = \frac{\overline{W}}{\tau} = f_{\max} k_B T. \quad (3.141)$$

Формулы (3.140) и (3.141) — очевидное следствие рассмотренной выше теоремы Найквиста.

Пропускная способность канала связи C (традиционное обозначение — идет от работ К. Шеннона) с учетом помех равна

$$C = K f_{\max} \ln \left(1 + \frac{P_s}{P} \right), \quad (3.142)$$

где K — постоянная, зависящая от выбора единиц;

P_s — мощность полезного сигнала, возможно, случайного;

P — мощность помех, определяемая (3.141).

Формула (3.142) — одна из наиболее известных универсальных теорем Шеннона безотносительно к способу передачи информации.

3.6.4 Теоремы Шеннона для сигналов с ограниченным спектром

Формула (3.142) — не единственная теорема Шеннона, полезная для выбора канала связи и/или проектирования аппаратуры. Более того, формулировка теорем в виде формул типа (3.142) — не единственная, даже не самая лучшая.

Пусть $n(t)$ — помехи,

$$n(t) = \sum_{l=-\infty}^{\infty} z_l \varphi_l(t) \quad (3.143)$$

— ортогональное разложение помех на интервале $-\frac{T}{2} < t < \frac{T}{2}$, где

$$\frac{1}{T} \int_{-T/2}^{+T/2} \varphi_l(t) \varphi_m(t) dt = \begin{cases} 1, & \text{если } l = m, \\ 0, & \text{в противном случае,} \end{cases}$$

а распределения величин z_l — гауссовы с нулевыми средними и дисперсиями $\sigma_z^2 = N_0/2T$, при этом $N_0\Delta F$ — средняя мощность, $P = N_0\Delta F$ в полосе частот ΔF , Гц.

Пусть в канале с помехами $n(t)$ передается полезный сигнал $s(t)$ со спектром, ограниченным заданной полосой $\Delta F = (0; f_{\max})$. Указанный сигнал складывается с помехами и образует смесь

$$v(t) = s(t) + n(t). \quad (3.144)$$

Сигналы v , s и n задаются отсчетами в $2T\Delta F$ -мерном пространстве, и формулу (3.144) удобно записать в векторном виде

$$\bar{v} = \bar{s} + \bar{n}.$$

Пусть $p(\bar{S})$ — плотность распределения вероятностей отсчетов $\bar{s}$. Имеем следующую полезную теорему Шеннона. Если

$$\int_{\bar{s} \in S} |\bar{S}|^2 p(\bar{s}) dS \leq P_s, \quad (3.145)$$

то пропускная способность

$$C = \Delta F \log \left(1 + \frac{P_s}{N_0 \Delta F} \right), \quad (3.146)$$

где ΔF — ширина полосы частот, Гц, а $N_0\Delta F = P$ — средняя в этой полосе мощность помех.

В этой теореме $K = 1$ и основание логарифма в (3.146) не фиксируется. Напомним, что введенная ранее пропускная способность канала определяется как

$$C = \max_{T; w(x,y)} \left[\frac{1}{T} I(X, Y) \right],$$

т. е. как максимум взаимной информации, вычисленной по всем значениям времени T и всем плотностям распределения вероятностей ансамблей.

Частным случаем этой теоремы при нулевой мощности помех $N_0 \Delta F = 0$ является еще одна теорема Котельникова:

$$C = f_{\max}. \quad (3.147)$$

Учитывая формулу Найквиста, можно записать

$$N_0 = kT, \quad (3.148)$$

где T — абсолютная температура канала.

Отсчетами сигнала на входе канала могут быть амплитудные отсчеты, взятые с временным шагом $\frac{1}{2f_{\max}}$ согласно изложенной выше теореме Котельникова.

3.6.5 Теоремы Шеннона для дискретных сообщений

Пожалуй, самая эффективная и полезная формулировка теоремы Шеннона дается в случае дискретных (импульсных) сигналов (сообщений). В дискретных ансамблях сигналов главным понятием служит символ — некое четко различимое физическое состояние, из комбинаций которого складываются информационные сообщения.

Теорема утверждает следующее: *если пропускная способность канала C , а количество информации, создаваемое передатчиком на входе канала, составляет i бит/символ в секунду, то существует способ кодирования, позволяющий передать по каналу в среднем*

$$\frac{C}{i} - \epsilon \quad (3.149)$$

символов в секунду, где ϵ — сколь угодно малая величина, $\epsilon \rightarrow 0$.

Самое удивительное в этой теореме, что она относится к каналу с помехами. Физический смысл ее предельно прост: если каждый символ несет в среднем i бит, то скорость передачи информации практически может быть доведена до пропускной способности C . Иными словами, всегда найдется код, который позволяет использовать канал связи почти с полной нагрузкой. И это в условиях помех!

Как найти этот код, другой вопрос. Его нужно решать отдельным приемом, который теорема указать не может. Указанному вопросу посвящены многочисленные работы по теории кодирования.

Дадим формулировку теоремы в терминах изложенной нами выше теории информации для дискретных сообщений.

3.6.6 Теорема Шеннона в терминах полезной информации

Для канала с помехами можно ввести характеристику

$$C_{\Pi} = \max \left[\frac{1}{T} \cdot I_{\text{полезн}}(X, Y) \right], \quad (3.150)$$

которую будет правильнее назвать пропускной способностью канала для не искаженной помехами информации. Максимум в (3.150) берется по всем возможным вероятностям генерирования сообщений передатчиком на входе АОКС. Ин-

формация $\mathbb{I}_{\text{полезн}}$ получается путем вычитания из полной взаимной информации $\mathbb{I}_{\text{вых}}(X, Y)$ ее помеховой составляющей $\mathbb{I}_{\text{помех}}(X, Y)$.

Пусть, как и выше, $\bar{p} = p_1, p_2, \dots, p_K$ — амплитудные вероятности появления сообщений $s_1, s_2, \dots, s_K$ на выходе передатчика или, что то же, на входе ОАК. Обозначим p_{jl} — переходные вероятности, которые при $j \neq l$ отвечают вероятностям помех. Тогда помеховая составляющая

$$\mathbb{I}_{\text{помех}}(X, Y) = - \sum_{j=1}^K \sum_{l=1}^K p_j p_{jl} \log_2(p_{jl}). \quad (3.151)$$

Общая информация в канале, т. е. информация на его выходе, составит:

$$\mathbb{I}_{\text{вых}}(X, Y) = - \sum_{j=1}^K q_j \log_2 q_j, \quad (3.152)$$

где q_j — выходные апостериорные вероятности сообщений $s_1, s_2, \dots, s_K$. Выразим апостериорные вероятности через априорные и переходные по формулам

$$q_j = \sum_{l=1}^K p_l p_{lj}, \quad (3.153)$$

подставим в (3.152) и выпишем выражение для $\mathbb{I}_{\text{полезн}}(X, Y)$.

Имеем

$$\begin{aligned} \mathbb{I}_{\text{полезн}}(X, Y) &= \mathbb{I}_{\text{вых}}(X, Y) - \mathbb{I}_{\text{помех}}(X, Y) = \\ &= - \sum_{j=1}^K \sum_{l=1}^K p_l p_{lj} \log_2 \left(\sum_{l=1}^K p_l p_{lj} \right) + \sum_{j=1}^K \sum_{l=1}^K p_j p_{jl} \log_2(p_{jl}). \end{aligned} \quad (3.154)$$

Теперь основная теорема Шеннона для канала с шумом.

Существует способ кодирования, позволяющий обеспечить передачу сообщения $\mathbb{S}$, состоящего из любого заданного количества M символов (при заданном количестве информации i на один символ в секунду, $i < C_n$), по каналу с помехами со сколь угодно малой вероятностью ошибочного приема. Если $i > C_n$, то существует способ кодирования, при котором потеря информации будет со сколь угодно близко приближаться к нижней границе $i - C$, но никогда не сможет стать меньше $i - C$.

Теорему, доказанную Шенноном, обобщали последователи, снимая предположения о дискретности, локальной ограниченности мощности сигнала, о гауссовости сигналов и «белизне» помех и т. д. В настоящее время теорема стала фундаментальным фактом теории связи. До доказательства этой замечательной теоремы теории связи фактически не было.

Теорема далеко не очевидна, можно сказать парадоксальна, и с трудом усваивается в инженерной практике.

Во-первых, роль помех сказывается в том, что, если количество передаваемой информации превысит пропускную способность, помехи все равно уничтожат любой «избыток» информации и сведут темп ее передачи к непревзойденному уровню C .

Во-вторых, часто кажется, что помехи должны непременно искажать любое сообщение в той или иной степени независимо от того, как оно закодировано. Даже если передавать сообщения с большой избыточностью, все равно невозможно, как считают некоторые, добиться нулевой вероятности искажения. Ведь найдется такое длинное сообщение, где эта вероятность неизбежно проявится, какой-нибудь символ да исказится.

Однако здесь интуиция нас обманывает! «Здравый» смысл ведет нас к неправильному выводу. Существуют «хитрые» способы кодирования, при которых искажения на выходе канала (в самом деле неизбежные, имеющие ненулевую вероятность!) удастся скорректировать и восстановить правильно все символы без исключения. Иногда почему-то забывают, что помеховые искажения — это не последние операции, которым подвергаются сообщения, что приемом искаженных помехами сообщений не заканчивается процесс передачи информации по каналу связи. Принятое сообщение никто не мешает подвергнуть на выходе дополнительной обработке. Именно она спасает положение — искаженное сообщение удастся *восстановить*, «очистить» от помех. Правда, восстановление может «дорого обойтись». Не исключено, что на него будет затрачено усилий больше, чем на *передачу* данных по каналу.

Теорема вдохновила и продолжает вдохновлять конструкторов аппаратуры на поиски «противопомеховых» способов кодирования. Пусть практическая их реализация часто сложна, громоздка, однако выгоды на выходе чаще всего превышают затраты на создание «антипомехового» оборудования.

ГЛАВА 4

Коды и их применение в системах передачи информации

В данной главе рассматриваются все виды кодирования информации в современных системах передачи дискретной информации. При этом любое кодирование есть преобразование одной цифровой информации в другую как на передающей стороне, так и на приемной. Однако цели у такого преобразования могут быть совершенно разные. Это либо устранение содержащейся в информации избыточности, либо, наоборот, сознательное внесение избыточности для повышения помехоустойчивости, либо шифрация информации.

4.1 Коды, устраняющие избыточность

4.1.1 Введение в теорию кодирования

Рассмотрим простейший пример.

Пусть имеется дискретный источник без памяти (ДИБП), на выходе которого может возникать одна из четырех букв с соответствующими вероятностями: a_1 , $P(a_1) = 0,5$; a_2 , $P(a_2) = 0,25$; a_3 и a_4 , $P(a_3) = P(a_4) = 0,125$. Если пронумеровать эти буквы последовательностями из двух бит равной длины, то будет тратиться в среднем два бита на букву, $N_{\text{ср}} = 2$.

Пронумеруем эти буквы другим способом: $a_1 \rightarrow 0$; $a_2 \rightarrow 10$; $a_3 \rightarrow 110$; $a_4 \rightarrow 111$. Тогда средняя длина буквы $N_{\text{ср}} = 1,75$, причем все буквы будут однозначно декодироваться, так как никакая более короткая буква не является префиксом (началом) более длинной буквы. Отметим, что энтропия данного источника

$$H = \sum_{i=1}^{i=K} -P(ai) \log_2 P(ai)$$

также равна 1,75. Отношение средней длины и длины буквы при равномерном и неравномерном кодировании может трактоваться как выигрыш или коэффициент сжатия.

Даже на таком простейшем примере видно, что сообщения содержат избыточность, устраняя которую можно «сжимать» информацию.

Предположим, что мы имеем дискретный канал, вероятность возникновения ошибки в котором близка к нулю (в идеале $= 0$). Такой канал называют идеальным каналом или каналом без шума. При наличии идеального канала естественно поставить вопрос о возможности передачи по нему без потерь информации от произвольного дискретного источника A , характеризуемого энтропией $H(A)$, со скоростью, равной пропускной способности канала. Блок-схема такой системы передачи информации показана на рис. 4.1.



Рис. 4.1. Блок-схема системы передачи информации через идеальный канал связи

Кодер осуществляет кодирование сообщений, т.е. каждому дискретному сообщению по определенному правилу ставит в соответствие последовательность символов из алфавита A . При этом по отношению к входу канала выдаваемые кодером символы сами являются дискретными элементами сообщений, статистические свойства которых должны отличаться от статистических свойств сообщений исходного источника. Возможность построения кодера, полностью устраняющего избыточность произвольного исходного источника сообщений, и определяет возможность решения поставленной задачи безошибочной передачи информации со скоростью, равной пропускной способности канала связи. При полном решении этой задачи энтропия как раз и оказывается равной средней длине буквы на выходе кодера.

Степень приближения к точному выполнению равенства зависит от степени уменьшения избыточности источника сообщений. Кодирование, позволяющее устранять избыточность источников сообщений, называется эффективным или статистическим. Коды, получаемые в результате такого кодирования, называются эффективными, статистическими или устраняющими избыточность.

Рассмотрим основные идеи, которые могут быть положены в основу эффективного кодирования, обусловленного двумя причинами:

- 1) памятью источника,
- 2) разной вероятностью исходных сообщений.

Универсальным способом уменьшения избыточности, обусловленной памятью источника, является укрупнение элементарных сообщений или букв. При этом кодирование осуществляется длинными блоками. Вероятностные связи между блоками меньше, чем между отдельными элементами сообщений, и чем длиннее блоки, тем больше эти вероятностные связи разрушаются. Смысл укрупнения можно пояснить на примере буквенного текста: если вероятностные связи между соседними буквами в любом языке достаточно сильны, то между словами они значительно слабее, еще меньше – между фразами, еще меньше – между абзацами. Поэтому, применяя кодирование слов, фраз, абзацев, можно практически

устранить избыточность, обусловленную вероятностными связями. Однако при этом возрастает задержка передачи сообщений и растет сложность кодирования и декодирования.

Уменьшение избыточности, обусловленной разной вероятностью исходных сообщений, может быть достигнуто применением неравномерных кодов. Основная идея построения таких кодов состоит в том, что наиболее вероятным сообщениям ставятся в соответствие наиболее короткие блоки кодовых символов (кодовые комбинации), а наименее вероятным — более длинные. В силу неравномерности кодов слов и случайного характера сообщения передача без потерь информации с постоянной скоростью возможна лишь при наличии буфера достаточной длины.

4.1.2 Теорема Шеннона для дискретного источника

Предельные возможности статистического кодирования раскрываются в теореме Шеннона для дискретного источника, что является одной из основ теории передачи информации. Эта теорема может быть сформулирована для кодовых слов одинаковой длины следующим образом [28].

Пусть дискретный источник без памяти имеет конечную энтропию $H(U)$. Рассмотрим кодирование последовательностей из L букв источника в последовательности их N кодовых букв, принадлежащих кодовому алфавиту объемом D . Каждой кодовой последовательности может сопоставляться только одна последовательность источника. Пусть P_e — вероятность появления последовательности источника, которой не соответствует никакая кодовая последовательность. Тогда, если при каком-либо $b > 0$

$$N/L \geq [H(U) + b]/\log D,$$

то величину P_e можно сделать произвольно малой, выбирая значение L достаточно большим.

Обратно, если

$$N/L \leq [H(U) - b]/\log D,$$

то P_e становится сколь угодно близкой к L , когда L становится достаточно большим.

Эта теорема имеет достаточно простое и эвристическое обоснование. Можно закодировать буквы произвольного дискретного источника без памяти таким образом, что энтропия кодовых букв будет максимальной.

В случае кодовых слов разной длины теорема имеет следующий вид [28].

Пусть источник сообщений имеет конечный ансамбль U и энтропию $H(U)$. Имеется кодовый алфавит из D символов. Можно так приписать кодовые слова буквам источника, что будет выполняться свойство префикса и средняя длина кодового слова N_{cp} будет удовлетворять условию

$$N_{cp} < H(U)/\log D + 1.$$

Более того, для любого однозначно декодируемого множества кодовых слов

$$N_{cp} \geq H(U)/\log D.$$

Полученный результат позволяет дать следующее толкование энтропии: энтропия источника есть наименьшее количество двоичных символов на сообщение

на выходе наилучшего кодера для этого источника при условии, что сообщения могут быть восстановлены по выходу кодера сколь угодно точно.

Естественно, возможно обобщение и на непрерывный случай.

Возможны различные варианты доказательства этой теоремы, иллюстрирующие два возможных подхода к построению эффективных кодов, основанных на использовании равномерного и неравномерного кодирования. При неравномерном кодировании обеспечивается однозначное декодирование всех сообщений.

Предложенный Шенноном метод эффективного кодирования практически совпадает с методом, предложенным другим американским ученым Фано, по которому сообщение длины L , записанное в порядке невозрастания вероятностей, разделяется на две части так, чтобы суммарные вероятности сообщений в каждой части были по возможности равны. Сообщениям первой части приписывается в качестве первого символа 0, сообщениям второй части 1. Затем каждая из этих частей (если она содержит более одного сообщения) опять делится на две примерно равные части, и в качестве второго символа для первой из них берется 0, а для второй 1. Этот процесс повторяется до тех пор, пока в каждой из полученных частей не останется по одному сообщению. Именно такой пример и был рассмотрен в начале раздела.

Существуют и другие методы эффективного кодирования. Кодирование по методу Шеннона-Фано, так же как и другими методами, может применяться не только к последовательностям из L -элементных сообщений, но и непосредственно к источникам неравновероятных элементарных сообщений. При этом уменьшается выигрыш в эффективности. В том случае, когда левая часть вышеприведенной системы неравенств обращается в равенство, имеем $N_{\text{ср}} = H(U)$. Код, обладающий этим равенством, называется оптимальным. Для того чтобы сообщение источника можно было закодировать двоичным оптимальным кодом, необходимо и достаточно, чтобы все вероятности источника сообщения представляли собой числа, равные целой отрицательной степени числа 2.

Рассуждая аналогичным образом, можно показать, что и в случае кодирования сообщений источника неравномерным кодом с произвольным основанием M оптимальный код может быть получен при условии равенства вероятности всех сообщений целым отрицательным степеням числа M . Если распределение вероятностей кодированного источника не обладает указанным свойством, эффективный код не будет оптимальным и соответствующая ему $N_{\text{ср}}$ будет больше энтропии $H(U)$. Величина $H(U)/N_{\text{ср}}$, характеризующая степень близости неравномерного статистического кода к оптимальному коду, называется эффективностью кода. Таким образом, нижний предел в условии теоремы может быть достигнут лишь при определенном распределении вероятности источника сообщений. Однако приближение к нему может быть сколь угодно близким при увеличении длины L последовательности кодируемых сообщений. При этом рост эффективности системы передачи информации сопровождается увеличением задержки сообщений.

4.1.3 Применение кодов, устраняющих избыточность

Наибольшее распространение получил способ построения эффективного кода, предложенный Хаффменом.

Рассмотрим его на примере. Пусть задан алфавит из пяти символов A_1 – A_5 и их вероятности. В табл. 4.1 наряду с этими исходными данными приведены также результаты кодирования по Хаффмену: кодовые слова и их длины N_i , а также средняя длина $N_{\text{ср}}$.

Таблица 4.1. Исходные данные и результаты кодирования по Хаффмену

A_i	P_i	Слова	N_i
A_1	0,25	10	2
A_2	0,17	00	2
A_3	0,08	010	3
A_4	0,35	11	2
A_5	0,15	011	3
$\sum P_i = 1,0$			$N_{\text{ср}} = 2,23$

Сама процедура построения кода Хаффмена проиллюстрирована в табл. 4.2.

Таблица 4.2. Процедура построения кода Хаффмена

A_i	P_i	Шаги объединения				Слова
		1	2	3	4	
A_1	0,35	0,35	0,40	0,60	1,00	11
A_2	0,25	0,25	0,35	0,40		10
A_3	0,17	0,23	0,25			00
A_4	0,15	0,17				011
A_5	0,08					010

На первом этапе символы упорядочивают по убыванию вероятностей, а затем выполняют несколько шагов «объединения», на каждом из которых суммируют вероятности наиболее редко встречающихся символов, и столбец вероятностей пересортировывается.

На втором этапе строится дерево кода, ветви которого отображают в обратном порядке процесс «объединения вероятностей». При построении дерева принимается правило соответствия большей вероятности одному из направлений ветви (например, «левому») и определенному значению бита кода, например 1. Цепочки битов от «корня» до конца каждой ветви соответствуют кодам исходных символов.

Процедура кодирования сводится к выбору из кодовой таблицы цепочек, соответствующих каждому символу источника. Декодирование предусматривает выделение в битовом потоке кодов символов и их расшифровку в соответствии с таблицей.

Код Хаффмена может быть двухпроходным и однопроходным. Первый строится по результатам подсчета частот (вероятностей) появления различных символов в данном сообщении. Второй использует готовую таблицу кодирования, построенную на основе вероятностей символов в сообщениях похожего типа. Например, кодирование текста на русском языке в первом случае включает его предварительный анализ, подсчет вероятностей символов, построение дерева кода и

таблицы кодирования индивидуально для данного сообщения. Во втором случае будет работать готовая таблица, построенная по результатам анализа множества русскоязычных текстов. Двухпроходной код более полно использует возможности сжатия. Однако при этом вместе с сообщением нужно передавать и кодовую таблицу. Однопроходной код не оптимален, однако прост в использовании, поэтому на практике обычно применяют именно его.

Рассмотренные выше основные предельные соотношения привели к широкому применению кодов, устраняющих избыточность во всех телекоммуникационных и вычислительных системах и сетях. Изложенный выше метод кодирования Хаффмена, хотя и является оптимальным, обладает существенными недостатками: большой задержкой буферирования, необходимостью знать или оценить исходные вероятности сообщений и согласованностью только с дискретным источником без памяти.

Здесь для компенсации первых двух недостатков нельзя не упомянуть простой и достаточно эффективный метод кодирования источника с неизвестным распределением вероятностей, известный как сжатие при помощи «стопки книг», или как сжатие сортировкой, или хешированием. Метод был разработан Рябко в 1980 году. Идея метода состоит в следующем: пусть алфавит источника состоит из K символов с номерами $1, 2, \dots, K$. Кодировующий алгоритм сохраняет последовательность символов, представляющую собой некоторую перестановку символов в последовательности первичного входного алфавита. При поступлении на вход некоторого символа, имеющего в этой переставленной последовательности номер i , кодировующий алгоритм записывает код этого символа (например, монотонный префиксный код). Затем поступивший символ переставляется в начало последовательности и номера всех символов, стоящих перед ним, увеличиваются на 1. Таким образом, наиболее часто встречающиеся символы будут переходить на 1. Таким образом, наиболее часто встречающиеся символы будут переходить в начало списка и иметь более короткие коды, что, в свою очередь, снизит объем выходного потока при их записи в качестве символов выходного потока.

Кодирование Лемпеля–Зива использует синтаксический метод для динамического источника и борется с третьей проблемой. Очевидно, что посимвольное кодирование, рассмотренное выше, не использует резервы сжатия информации, связанные с повторяемостью цепочек символов, т.е. памятью канала. Наиболее удачным алгоритмом сжатия, основанным на таком подходе, является алгоритм Лемпеля–Зива, который в разных модификациях используется, в частности, в большинстве программ-архиваторов. Основная идея алгоритма состоит в том, что цепочки символов, уже встреченные ранее, кодируются ссылкой на их «координаты» (номер первого символа и длину) в «словаре», где находится уже обработанная часть сообщения. Более детально основные идеи алгоритма иллюстрирует рис. 4.2.

«Сжимаемое» сообщение постепенно входит в буфер источника. Ядро кодера выделяет в буфере блок (цепочку) символов первоначально максимальной длины (обычно порядка 16 символов) и пытается найти совпадающую цепочку в словаре источника. Если это не удастся, кодер повторяет поиск для более короткого «урезанного» варианта цепочки. Когда эта цепочка обнаруживается в словаре, в канал передаются ее координаты. Если же поиск не дал результата даже для самого короткого варианта цепочки из двух символов, каждый из них передается по каналу самостоятельно.

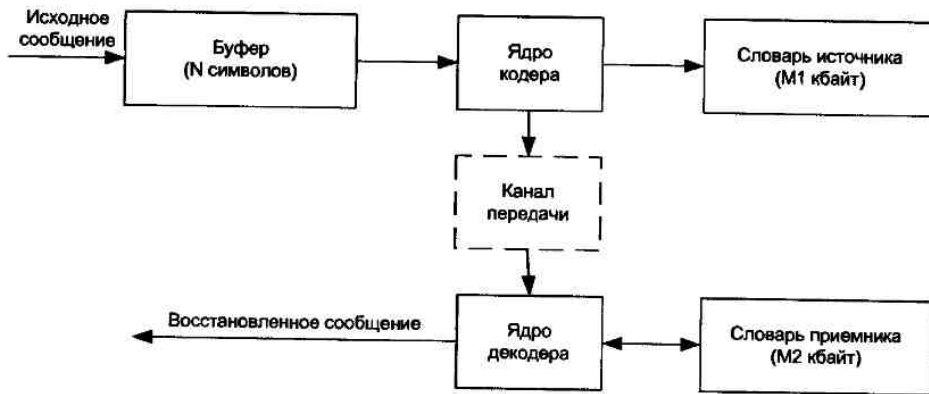


Рис. 4.2. Иллюстрация идеи алгоритма Лемпеля–Зива

На приемной стороне ядро декодера принимает коды и восстанавливает исходное сообщение по собственному словарю. При этом восстановленные цепочки тут же попадают в словарь приемника так, что его содержимое синхронизируется с содержимым словаря источника.

При этом следует помнить, что:

- коды координат цепочки и коды отдельных символов различаются битовыми признаками (например, в первом случае — 1, во втором — 0);
- поскольку цепочки находятся чаще в начале словаря и чаще бывают короткими, дополнительный выигрыш получают за счет статистического кодирования (по Хаффмену) их «адресов» и «длин»;
- «канал» — понятие, применимое и к реальному каналу передачи данных, и к файлу, куда данные записываются для хранения. В последнем случае декодер «отрабатывает» при разворачивании сжатого файла;
- при ограниченной длине словаря (обычно от 4 до 16 кбайт) новые поступающие символы и цепочки «вытесняют» прежние (текст как бы «вдвигается» в словарь). Разумеется, вначале, когда словарь не заполнен, эффективность сжатия невысока. Рост объема словаря позволяет повысить степень сжатия, но значительно увеличивается трудоемкость поиска цепочек.

Алгоритм Лемпеля–Зива используется в большинстве популярных программ-архиваторов (в том числе, например, в zip, rar, arj и их windows-версиях). Различия скорости и эффективности операций кодирование-декодирование определяются в основном особенностями программной реализации. Алгоритм Лемпеля–Зива требует большого количества вычислительной работы. Его модификация — алгоритм Лемпеля–Зива–Велча является менее трудоемким, хотя и дает несколько худшие результаты по сжатию.

На похожих принципах и основаны все существующие методы сжатия одномерных цифровых файлов (архиваторы) или одномерных цифровых потоков

(например, оцифрованная речь). При оцифровке речи, правда, возникает дополнительная специфика.

Оцифровка звука сводится к дискретизации (взятию отсчетов U_i амплитуды сигнала с шагом по времени Δt или частотой $f_g = \frac{1}{\Delta t}$), последующему квантованию с шагом ΔU и аналого-цифровому преобразованию.

С учетом особенностей слухового аппарата человека и принимая во внимания теорему Котельникова, стандартизированы следующие уровни частот:

- 44 кГц (для тренированного уха максимальной воспринимаемой считается $f_m = 20$ кГц);
- 22 кГц (в случае, когда восприятие высокочастотных составляющих не критично $f_m = 10$ кГц);
- 8 кГц (для речевого сигнала — с учетом экспериментально установленного порога «слоговой разборчивости» $f_m = 3,4$ кГц).

Способ представления непрерывного сигнала последовательностью кодов, которые отражают амплитуду импульсов-отсчетов, называется импульсно-кодовой модуляцией (ИКМ).

Частота дискретизации согласно теореме Найквиста–Котельникова должна быть не меньше удвоенной максимальной частоты спектра сигнала.

Погрешности квантования определяются исходя из емкости кода для максимального уровня сигнала — 8 или 16 бит на отсчет.

Один из классических способов сжатия — так называемая «адаптивная дифференциальная ИКМ» (АДИКМ). Она основана на следующем:

- вместо кода абсолютного значения очередного импульса передают код изменения его амплитуды по отношению к предыдущему (дифференциальному);
- при увеличении скорости передачи сигнала шаг его квантования возрастает, а при уменьшении — уменьшается.

Использование АДИКМ позволяет уменьшить объем передаваемых данных примерно на порядок.

Второй способ сжатия именуется компадированием и использует известную особенность человеческого слухового аппарата, вследствие которой ухо менее критично к погрешностям квантования при более громком звуке.

Механизм компадирования работает следующим образом. Звуковой сигнал перед оцифровкой подается на так называемый «компрессор» КМП — нелинейный преобразователь, «растягивающий» слабые сигналы и «сжимающий» сильные. После передачи зашумленный погрешностями АЦП сигнал подается на вход «эспандера», который в противоположность компрессору сжимает слабые сигналы и растягивает сильные, восстанавливая первоначальный баланс, при этом погрешности ΔU , попавшие в область слабых сигналов, уменьшаются, а сильных — возрастают.

Это не только позволяет учесть особенности человеческого восприятия звука, но и приводит к уменьшению «среднестатистических» погрешностей квантования — дело в том, что слабые сигналы встречаются чаще.

Стоит добавить, что в современных системах вместо комбинации аналоговый компрессор (эспандер) + АЦП (ЦАП) реально используются АЦП с нелинейной характеристикой.

Наряду с описанными, при сжатии звука используются еще дополнительно некоторые способы, например классическое сжатие.

Однако часто цифровые файлы, или потоки, бывают двумерными, например возникающими в результате оцифровки изображений. В этом случае использование двумерных (строчных и столбцовых) вероятностных связей при сжатии изображений дает дополнительный эффект. В настоящее время в мире существуют три стандарта представления цветного телевизионного видеосигнала: PAL, SECAM и NTSC. Все они преобразуют при оцифровке исходный аналоговый видеосигнал в форму, удобную для передачи по линии связи. Несмотря на различие в методе кодирования видеoinформации, все три системы телевидения представляют исходный телевизионный сигнал в виде сигнала яркости (Y) и двух цветоразностных ($R - Y$) и ($B - Y$) сигналов, а затем каждая по-своему кодирует его.

Один кадр телевизионного изображения представляется в цифровом виде как поле 768×576 точек или, как еще говорят, пикселей — в системе PAL/SECAM и 640×480 — в NTSC. Так как человеческий глаз более чувствителен к перепадам яркости, чем к цвету, важнее точная передача сигнала яркости, при оцифровке видеосигнала соотношение числа отсчетов сигнала яркости и цветоразностных сигналов принято равным $4 : 2 : 2$, т. е. частота измерения сигнала Y в два раза выше, чем частота измерения сигналов $R - Y$ и $B - Y$, — одна строка оцифрованного телевизионного видеосигнала содержит 768 измерений яркости и по 384 измерения каждого из цветоразностных сигналов. Соотношение частот оцифровки яркости и цвета также определяет качество изображения, которое впоследствии может быть восстановлено. Это отношение может выражаться соотношениями от $4 : 1 : 1$ для формата DV до $4 : 4 : 4$ для Digital формата. Оцифровка по стандарту $4 : 4 : 4$ предполагает использование одной и той же частоты дискретизации для яркостного и цветоразностных сигналов. Выбор такого способа оцифровки очень хорошо стыкуется с системами PAL/SECAM и NTSC, разница заключена в количестве строк и кадров в секунду ($576/25$ для PAL/SECAM и $486/30$ для NTSC).

Нетрудно подсчитать, что при использовании 10-битного АЦП одна секунда оцифрованной видеозаписи будет занимать на диске файл размером 27 684 000 байт, а для хранения всего 49 секунд видеoinформации со звуком потребуется более 1 Гбайт свободного дискового пространства (что соответствует скорости передачи более 160 Мбит/с).

Чтобы уменьшить размеры файла или скорости цифровых потоков с видеoinформацией, применяют различные методы сжатия (кодеки). Наиболее распространенными являются два метода сжатия: M-JPEG (неподвижная «картинка») и MPEG (движущееся изображение). Оба метода основаны на так называемом дискретно-косинусном преобразовании изображения (частный случай дискретного преобразования Фурье), разработанном для формата JPEG. Отличие состоит в том, что при сжатии по методу M-JPEG устраняется внутрикадровая избыточность, т. е. такое сжатие устраняет избыточность каждого кадра в отдельности. Сжатие по методу MPEG, кроме устранения внутрикадровой избыточности, устраняет и межкадровую.

Метод MPEG имеет много разновидностей, например MPEG1 и MPEG2. Метод MPEG1 не поддерживает чересстрочной развертки и работает только с половинным разрешением. Для сохранения оцифрованного видеосигнала с полным разрешением нужно использовать формат MPEG2. Следует отметить, что MPEG-сжатие существенно сложнее M-JPEG и требует специального оборудования или программного обеспечения. Существующие сегодня алгоритмы и их реализации достигают коэффициентов сжатия в 30 раз и более без заметного ухудшения качества изображения.

Приведем краткое описание алгоритма JPEG.

В настоящее время широко используется «блочная» версия алгоритма, в которой все изображение разбивается на блоки 8×8 и в дальнейшем эти блоки «огрубляются» таким образом, чтобы код, который их описывает, стал как можно короче (исходное описание каждого такого блока требует $8 \times 8 \times 3 = 192$ байта). Дальнейший алгоритм включает следующие этапы.

- Происходит определенная взаимобратимая смена кодирования от RGB к YUV.
- Выделяются блоки 8×8 пикселей. При этом каждой из составляющих YUV отвечает матрица коэффициентов P_{ij} . (Например, в матрице U значение $P_{ij} = 0$ означает, что у данной точки отсутствует красная цветовая составляющая, а $P_{ij} = 255$ значит, что она будет максимально яркой.)
- Выполняется так называемое «прореживание». Четверки блоков изображения 8×8 объединяются в «макроблоки» 16×16 , а затем для цветковых составляющих U и V из соответствующих матриц исключаются все четные строки и столбцы. При этом матрица яркостной составляющей Y остается нетронутой. В итоге количество коэффициентов сокращается вдвое (вместо $4 \times 3 = 12$ блоков остается 6).
- Для всех оставшихся блоков изображения выполняется так называемое дискретное косинус-преобразование (ДКП). При этом матрицы $N \times N$ ($N = 8$) коэффициентов P преобразуются в матрицы D в соответствии со следующей процедурой:

$$d_{ij} = \frac{1}{2N} C_i C_j \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} p_{xy} \cos\left[\frac{(2x+1)i\pi}{2N}\right] \cos\left[\frac{(2y+1)j\pi}{2N}\right].$$

Смысл косинус-преобразования заключается в том, что коэффициенты d_{ij} отражают «амплитуду колебаний» яркости пикселей. Например, если все пиксели блока имеют одинаковую яркость, то максимальными будет коэффициент d_{11} , а остальные $d_{ij} = 0$. Чем больше деталей в изображении, тем большими будут значения «удаленных» коэффициентов.

- Все коэффициенты масштабируются, т.е. огрубляются. Если их поделить, скажем, на 8, то длина кода для каждого из них сократится на 3 бита из 8. Но поскольку «удаленные» коэффициенты (правая нижняя часть матрицы) обычно малы, они просто обнуляются в результате масштабирования.

Нужно сказать, что именно на этом этапе информация о деталях изображения необратимо теряется. Регулируя величину делителя, можно задавать соотношение «степень сжатия—качество восстановленного изображения».

- Последний этап — кодирование оставшихся коэффициентов. Он включает несколько шагов: коэффициенты записываются в цепочку в порядке «обхода» (в примере цепочка имеет вид: 16 8 6 1 3 4 0 0 1 ...); последовательности нулей в цепочке кодируются методом повторов; длины «нулевых» серий кодируются по Хаффману. В результате подобного кодирования из примерно полутора тысяч бит, описывающих блок, остается обычно несколько десятков — код изображения сжимается в десятки раз.

4.2 Общее понятие о шифровании информации

Рассмотрим классическую модель Шеннона для канала связи с шифрованием. Между источником и приемником существует два канала — открытый и закрытый. В открытом канале передачи K_n° сообщение T от источника I либо может быть перехвачено «злоумышленником» $Зл$, который получает несанкционированный доступ к его содержанию, либо $Зл$ может сообщение исказить, либо $Зл$ может подставить свое сообщение вместо сообщения источника. Первая задача соответствует классическому шифрованию, вторая — работе в условиях преднамеренных помех, третья — электронной подписи. Рассмотрим для простоты первую задачу.

Чтобы избежать перехвата, кодер K шифрует сообщение с помощью функции $Ш = F_{ш}(T_1 K)$, которая использует секретный ключ K . Декодер $ДК^ш$ выполняет обратную операцию дешифрования $T = F_{от}(Ш_1 K)$, применяя тот же ключ K . Чтобы доставить ключ K получателю сообщений, используется секретный канал $K^с$, в котором информация не может быть перехвачена злоумышленником. Разумеется, организовать передачу по секретному каналу намного сложнее, зато, один раз доставив с его помощью секретный ключ, можно затем многократно использовать этот ключ, передавая зашифрованные данные по открытому каналу.

Одним из первых исторических примеров, реализующих эту схему, был так называемый шифр Цезаря, который предполагал замену всех букв сообщения на буквы циклически сдвинутого алфавита. Например, если циклический сдвиг латинского алфавита выполнить на $n = 3$ позиции, то букве a будет соответствовать d , $c — e$ и т. д. В данном случае ключом является значение сдвига n . Зная сам алгоритм подстановки, подобрать ключ не так сложно (перебор по длине алфавита). Между тем теория шифрования исходит из того факта, что сам алгоритм шифрования всегда может быть известен, а неизвестен только ключ. Разумеется, для этого ключ должен быть достаточно длинным, чтобы количество его возможных вариантов было велико, а функция шифрования/дешифрования — достаточно сложной, чтобы затруднить прямой перебор. Например, если вернуться к шифру Цезаря, можно предложить усовершенствованный способ подстановки, когда исходному алфавиту соответствует измененный, который и является

ключом

$$\begin{array}{l} abcde \dots xyz \\ foylnl \dots rem. \end{array}$$

Теперь количество вариантов ключа стало огромным (порядок 10^{26}). Однако, если использовать сведения о частоте появления отдельных букв, выделять повторяющиеся слова и т. д., перебор становится существенно меньше.

Интересно, что теоретически нераскрываемый шифр все же существует. Чтобы обеспечить абсолютную защиту, необходимо использовать ключ, длина которого не меньше длины сообщения, и применять этот ключ только один раз.

Существуют разные классы шифров с секретным ключом. Блочные шифры предполагают разбиение всего сообщения T на стандартные блоки и шифрование каждого блока

$$Ш_i = F(T_i, K).$$

Такой подход ускоряет решение задачи. Вместе с тем его слабым местом считается повторяемость блоков, которая упрощает несанкционированный доступ.

Примеры шифрования подстановкой рассмотрены на примере шифра Цезаря. Возможны также шифры на основе перестановок, которые дополнительно затрудняют несанкционированный доступ. Идея состоит в записи сообщения в некоторую матрицу по столбцам, а считывании по строкам, причем в порядке следования букв в слове-ключе в алфавите. На практике используют сочетания подстановки и перестановки — комбинированные блочные шифры.

Возможны также непрерывные или поточные шифры, в которых сообщение обрабатывается как единое целое. Наиболее распространенный способ — так называемое «гаммирование». «Гаммой» называют битовую последовательность Γ , которая накладывается на исходное сообщение, представленное в двоичном виде (используется поразрядное суммирование по модулю 2)

$$Ш = T \oplus \Gamma.$$

Сама гамма — это так называемая псевдослучайная последовательность двоичных чисел, получаемая, например, с помощью формулы

$$C_{i+1} = (A + C_i + B) \bmod M,$$

где A и B — константы, M — основание модуля. При такой процедуре ключом K является стартовое число C_0 . Достоинство такого способа шифрования — его высокое быстродействие. Его слабость в том, что, когда имеется уже зашифрованный текст, по нему легко восстановить участки гаммы, т. е. низкая устойчивость. Способ гаммирования в ряде случаев используют совместно с элементами блочного шифрования.

Наибольшее распространение в прошлом веке получил алгоритм шифрования DES (Data Encrypting Stand), разработанный в 1977 году компанией IBM и принятый как стандарт в США, а затем и во всем мире.

Стандарт DES реализует блочное шифрование данных с использованием 56-битного ключа. Каждый 64-битный блок данных T_i разделяется на «левую» и «правую» половины A_i и B_i . Шифрование включает 16 итераций перестановок и подстановок, в ходе которых к преобразуемым A_i и B_i «подмешиваются» раз-

личные фрагменты ключа K (56 бит ключа являются секретными, а еще 8 бит — проверочными символами). Операции шифрования и дешифрования были достаточно удачно оптимизированы по скорости, а надежность шифра уже почти четверть века позволяет ему оставаться стандартом. Хотя сейчас количество вариантов ключа 2^{56} уже считается недостаточно большим (в принципе этот недостаток можно обойти за счет, например, двукратного шифрования на ключах K_1 и K_2 , тогда криптостойкость определяется их суммарной длиной).

Однако с появлением и развитием Интернета на повестку дня стала задача передавать секретный ключ по тому же каналу, что и сообщение. Действительно, при сотнях миллионов пользователей Интернета невозможно организовать такое же количество секретных каналов. Тогда были предложены системы с открытыми ключами. Идея таких систем состоит в следующем. У каждого пользователя сети имеется два ключа — открытый K_o и закрытый K_z . Ключи K_o и K_z генерируются по определенному алгоритму самим абонентом, но получить один ключ из другого невозможно. Открытые ключи всех абонентов публикуются в сети, а закрытые абоненты хранят у себя.

В случае передачи зашифрованной информации от A к B абонент A закрывает информацию открытым ключом абонента B — $K_{об}$ по алгоритму $\Pi = F_{\Pi}(T_1 K_{об})$. Абонент B раскрывает информацию своим закрытым ключом $K_{зб}$ по алгоритму $T = F_{от}(\Pi_1 K_{зб})$.

В случае передачи информации с электронной подписью задача инверсна. Абонент A подписывает свою информацию с использованием своего закрытого ключа $K_{за}$ по алгоритму $\Pi = F_{\Pi}(T_1 K_{за})$, а любой другой абонент сети может убедиться в том, что информация подписана A при помощи открытого ключа A — $K_{оа}$ по алгоритму $T = F_{от}(\Pi_1 K_{зб})$. Система с открытыми ключами является чрезвычайно распространенной.

4.3 Корректирующие коды

4.3.1 Блок-схема системы связи и примеры простейших кодов

На рис. 4.3 показана блок-схема системы связи в случае наличия кодера и декодера корректирующего кода, при этом мы считаем, что кодирование, устраняющее избыточность, включено в источник и получатель дискретной информации. Также мы считаем, что канал связи дискретный по входу и по выходу, т. е. модулятор и демодулятор включен в канал. Следует отметить, что мы постоянно будем пользоваться искусственными приемами укрупнения тех или иных блоков теоретико-информационных блок-схем — каналов, источников/получателей информации и т. д.

Рассмотрим два простейших примера корректирующего кода.

Код с проверкой на четность

Список кодовых слов, каждое из которых состоит из информационной и проверочной части, показан ниже (табл. 4.3).



Рис. 4.3. Блок-схема системы связи при наличии кодера и декодера корректирующего кода

Таблица 4.3. Список кодовых слов

Информационная часть	Проверочная часть
000	0
001	1
010	1
011	0
100	1
101	0
110	0
111	1

Как мы видим, проверочный символ равен сумме по модулю 2 всех трех информационных. Легко также заметить, что кодовые слова отличаются друг от друга не менее чем в двух позициях. На три информационных символа кода тратится один проверочный, т. е. относительная скорость кода равна $3/4$, а избыточность $1/4$.

Пусть кодовые слова передаются побитно по двоичному симметричному каналу (ДСК), где p — вероятность ошибки, а $(1 - p)$ — правильной передачи. Легко заметить, что, если в кодовом слове исказился один символ или три, эта ситуация может быть обнаружена общей проверкой на четность (определением принадлежности кодового слова к коду). Если же произошло два или четыре искажения (ошибки), слово будет считаться правильным, т. е. произойдет неисправимая ошибка.

Таким образом, возможны три результата декодирования.

A — правильное декодирование (нет ошибок в канале);

B — обнаружение ошибки (одна или три ошибки в слове);

V — ошибочное декодирование (две или четыре ошибки в слове).

Соответственно, вероятности этих событий вычисляются следующим образом.

$$P_A = (1 - p)^4; \quad P_B = 4p(1 - p)^3 + 4p^3(1 - p); \quad P_V = 6p^2(1 - p)^2 + p^4.$$

Если, например, вероятность ошибки в канале ДСК равна 10^{-3} , то вероятность ошибки декодирования равна примерно $6 \cdot 10^{-6}$, а вероятность обнаружения ошибки $4 \cdot 10^{-3}$. С ухудшением канала (стремлением p к $1/2$) вероятность ошибки де-

кодирования P_B растет, но стремится к $7/16$, а вероятность обнаружения ошибки P_E также растет, но стремится к $3/8$. Вероятность правильного декодирования P_A падает и стремится к $1/16$. Естественно, этот код имеет хоть какой-либо смысл при малых вероятностях ошибки p в канале.

Код-повторение

Список, состоящий всего из двух кодовых слов, показан ниже (табл. 4.4).

Таблица 4.4. Список кодовых слов

Информационная часть	Проверочная часть
0	000
1	111

Как мы видим, проверочные символы повторяют один информационный. Легко также заметить, что кодовые слова отличаются друг от друга в четырех позициях. На один информационный символ кода тратится три проверочных, т. е. относительная скорость кода равна $1/4$, а избыточность $3/4$.

Можно предложить алгоритм декодирования голосованием по большинству одинаковых символов. Тогда правильное декодирование будет происходить при одной или менее ошибок в канале, ошибочное декодирование — при трех или более. При двух ошибках в канале будет обнаруживаться ошибка.

$$P_A = (1-p)^4 + 4p(1-p)^3; \quad P_E = 6p^2(1-p)^2; \quad P_B = 4p^3(1-p) + p^4.$$

Если, например, вероятность ошибки в канале ДСК равна 10^{-3} , то вероятность ошибки декодирования равна примерно 4×10^{-9} , а вероятность обнаружения ошибки 6×10^{-6} . С ухудшением канала (стремлением p к $1/2$) вероятность ошибки декодирования P_B растет, но стремится к $5/16$, а вероятность обнаружения ошибки P_E также растет, но стремится к $3/8$. Вероятность правильного декодирования P_A падает и стремится к $5/16$. Если сравнивать этот код с предыдущим, видно, что у него существенно выше помехоустойчивость, правда, достигается это ценой существенно большей избыточности. Таким образом, даже на таком простом примере видно, что, как для любой системы связи, существует обмен между помехоустойчивостью и скоростью передачи.

Также можно изменением алгоритма декодирования (порога принятия решения) «перекачивать» вероятности событий A , B и V . Повысим порог при декодировании второго кода — будем принимать решение, только если 4 символа будут равны 0 или 1. В противном случае всегда будет обнаруживаться ошибка. Тогда вероятности принимают следующий вид:

$$P_A = (1-p)^4; \quad P_E = 4p(1-p)^3 + 6p^2(1-p)^2 + 4p^3(1-p); \quad P_B = p^4.$$

Мы видим, что существенно уменьшилась вероятность ошибки декодирования, правда, ценой увеличения вероятности обнаружения ошибки и уменьшения вероятности правильного декодирования.

Таким образом, на простейших кодах мы проиллюстрировали ключевые свойства теории корректирующего кодирования. Для достижения пропускной спо-

способности канала необходимо увеличивать длину кодов. В данном случае код с проверкой на четность будет практически всегда обнаруживать ошибки, а скорость кода повторения будет стремиться к нулю, т.е. ни тот ни другой код не интересны для практического использования, однако являются хорошей иллюстрацией поведения реальных кодов. Теория корректирующего кодирования, начиная с первых работ Шеннона, как раз и создавала такие легко кодируемые и декодируемые коды на больших длинах, с помощью которых можно стремиться к пропускной способности канала при фиксированной и ненулевой скорости передачи. Для короткого обзора таких методов нам понадобятся некоторые математические определения.

4.3.2 Теорема Шеннона для канала с шумами

Данная теорема является фундаментальным положением теории информации и называется также основной теоремой кодирования Шеннона. Она может быть сформулирована следующим образом.

Если энтропия источника сообщений $H(U)$ меньше пропускной способности канала C ($H(U) < C$), то существует такая система кодирования, которая обеспечивает возможность передачи сообщений источника со сколь угодно малой вероятностью ошибки (или со сколь угодно малой ненадежностью). Если $H(U) > C$, то можно закодировать сообщение таким образом, что ненадежность в единицу времени будет меньше, чем $H(U) - C + \epsilon$, где ϵ стремится к нулю (прямая теорема).

Не существует способа кодирования, обеспечивающего ненадежность в единицу времени меньшую, чем $H(U) - C$ (обратная теорема).

В такой формулировке эта теорема была представлена Шенноном [111]. В литературе часто вторая часть прямой теоремы и обратная теорема объединяются в виде обратной теоремы, сформулированной так: если $H(U) > C$, то такого способа кодирования не существует.

Для доказательства первой части прямой теоремы используется множество длинных последовательностей элементарных дискретных сообщений источника длиной T , распадающееся на подмножества высоковероятных или типичных и маловероятных или нетипичных последовательностей. Пусть при некотором ансамбле входных сигналов дискретного канала XU обеспечивается пропускная способность канала

$$C = \max_{Q(0), Q(1), \dots, Q(K-1)} I(X, Y) = \\ = \max_{Q(0), Q(1), \dots, Q(K-1)} \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} Q(k) P(j) \log \left[P(j/k) / \sum_{i=0}^{K-1} Q(i) P(j/i) \right],$$

где $Q(0), Q(1), \dots, Q(K-1)$ — вероятности возникновения K букв на входе канала, $P(j/k)$, $k = 0, 1, \dots, K-1$; $j = 0, 1, \dots, J-1$ — переходные вероятности для дискретного канала без памяти с K входами и J выходами, $I(X, Y)$ — средняя взаимная информация между входом X и выходом Y дискретного канала без памяти.

В процессе кодирования каждой типичной последовательности источника ставится в соответствие одна из типичных последовательностей канальных сигналов. Нетипичные последовательности сообщений длительности T (если источник все же выдаст одну из них) не передаются, соглашаясь с тем, что каждая такая последовательность будет принята ошибочно. После выполнения указанного кодирования всеми возможными случайными способами проводится усреднение вероятности ошибок по всему этому большому классу возможных систем кодирования. Это равносильно вычислению вероятности ошибок при случайном связывании типичных последовательностей источника сообщений и канальных сигналов. Далее при помощи усреднения можно оценить среднюю вероятность ошибки.

При любом заданном $\epsilon > 0$ можно выбрать столь большое T , что будем иметь среднюю вероятность ошибки меньше ϵ . Но если среднее некоторого множества чисел не больше, чем ϵ , то в этом множестве должно существовать по крайней мере одно число меньше ϵ . Тогда среди всех возможных M кодов, обеспечивающих среднее значение вероятности ошибки, обязательно существует хотя бы один, у которого вероятность ошибки не превышает среднее. Таким образом, доказывается первая часть теоремы.

Вторую часть прямой теоремы легко доказать исходя из того, что можно просто передавать C бит в секунду от источника сообщений, совсем пренебрегая остатком создаваемой информации.

Физический смысл эффекта повышения вероятности при увеличении длительности кодируемых сообщений, вытекающий из доказательства прямой теоремы, заключается в том, что с ростом T увеличивается степень усреднения шума, действующего в канале, и, следовательно, уменьшается степень его мешающего воздействия. Кодирование сообщений длительности T способом, предполагаемым при доказательстве теоремы Шеннона, может начаться лишь тогда, когда сообщение целиком поступило на кодирующее устройство. Декодирование же может начаться, когда вся принятая последовательность поступила на декодирующее устройство. Поэтому суммарная задержка сообщений во времени между входом кодера и выходом декодера есть $2T + T_0$, где T_0 — время, затрачиваемое на кодирование, декодирование и прохождение по каналу. При большом T можно принять, что суммарная задержка есть $2T$. Из теоремы также следует принципиальная возможность обмена между вероятностью ошибки, задержкой и скоростью передачи информации. На практике сложность кодирования и декодирования существенно возрастает с ростом T .

4.3.3 Введение в теорию групп, колец и полей

Приведем короткую сводку теории групп, колец и полей в минимальном объеме, необходимом для чтения литературы по корректирующим кодам.

Группы

Алгебраическая система $\{G, O\}$, образованная непустым множеством G и операцией O , определенной для любых двух элементов a и b из G , называется группой, если выполнены следующие четыре аксиомы.

A1 (замкнутость).

Для любых двух элементов a и b из G однозначно определен элемент aOb , принадлежащий G .

A2 (ассоциативность).

Для любых трех элементов a, b и c из G выполняется равенство

$$aO(bOc) = (aOb)Oc.$$

A3 (существование единичного элемента).

В G существует элемент e , называемый единичным, такой, что для любого элемента a из G

$$aOe = eOa = a.$$

A4 (существование обратного элемента).

Для любого элемента a из G существует элемент x , называемый обратным к a , такой, что

$$xOa = aOx = e.$$

Группа $\{G, O\}$ называется коммутативной или абелевой, если справедлива следующая аксиома.

A5 (коммутативность).

Для двух произвольных элементов a и b из G

$$aOb = bOa.$$

Если в качестве групповой операции используется $+$, т.е. $\{G, +\}$, то группа называется аддитивной.

Группами является множество целых чисел с операцией обычного сложения и множество комплексных чисел с соответствующим сложением.

Число элементов в группе G называется порядком G . Группа, порядок которой конечен, называется конечной.

Легко доказывается, что единичный элемент в группе единственный и что для каждого элемента группы обратный элемент также единственный.

Пусть $\{G, O\}$ группа, а H — подмножество G , являющееся группой относительно той же групповой операции O . Тогда H называется подгруппой G .

Пусть H подгруппа группы G , а g — произвольный элемент из G . Множество элементов x , таких, что $x = gh$ (для всех h , принадлежащих H), обозначенное ниже через gH , называется левым смежным классом группы G по подгруппе H , порожденным элементом g . Аналогично можно ввести понятие правого смежного класса.

Можно доказать (Лагранж), что порядок любой подгруппы конечной группы является делителем порядка группы.

Тогда любую конечную группу можно однозначно разложить на смежные классы. Если M — порядок конечной группы, N — порядок ее подгруппы, $M = NJ$, то элементы группы могут быть разложены в двумерную таблицу размера J

строк на N столбцов, где первая строка есть собственно элементы подгруппы, а остальные строчки есть ее смежные классы.

Более подробно с теорией групп можно ознакомиться в монографиях по алгебраической теории кодирования [6, 41, 53, 59, 77].

Кольца

Рассматриваемые ниже кольца и поля в отличие от групп являются алгебраическими системами с двумя операциями.

Пусть R — множество, для любых двух элементов которого определены две операции $+$ и $*$. Алгебраическая система $\{R, +, *\}$ называется кольцом, если выполняются следующие три аксиомы.

Б1. $\{R, +\}$ — коммутативная группа.

Б2. $\{R, *\}$ — полугруппа, т. е. алгебраическая система, для которой выполняются аксиомы А1 и А2.

Б3 (дистрибутивность). Для любых трех элементов a, b и c из R выполняются тождества

$$\begin{aligned}a(b + c) &= ab + ac, \\(a + b)c &= ac + bc.\end{aligned}$$

Если операция $*$ коммутативна в кольце, то кольцо называется коммутативным. Если в кольце существует единичный элемент относительно операции $*$, то кольцо называется кольцом с единицей.

Целые числа с обычным сложением и умножением являются, например, коммутативным кольцом с единицей.

Множество всех многочленов вида

$$a_0x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n, a_i \in R, 0 \leq n$$

с коэффициентами из R является кольцом относительно обычных операций сложения и умножения многочленов. Это кольцо называется кольцом многочленов от неизвестного x над кольцом R и обозначается через $R[x]$.

Подмножество I кольца R называется идеалом кольца R , если оно удовлетворяет следующим двум условиям.

В1. Для любых двух элементов a и b из I элемент $a + b$ принадлежит I .

В2. Для любого элемента x кольца R и любого a из I оба произведения ax и xa лежат в I .

Если I идеал кольца R , то в R следующим образом можно ввести новое бинарное отношение

$$a \equiv b \pmod{I} \Leftrightarrow a - b \in I \Leftrightarrow a \equiv b \pmod{I} \Leftrightarrow a - b \in I.$$

Поля

Пусть F — множество, содержащее, по крайней мере 2 элемента, такое, что для любых двух элементов a и b из F определены две операции $+$ и $*$. Алгебраическая система $\{F, +, *\}$ называется полем, если она удовлетворяет следующим аксиомам.

Г1. $\{F, +\}$ — коммутативная группа.

Пусть O — единичный элемент группы $\{F, +\}$, а F^* — множество, полученное из F удалением элемента O .

Г2. $\{F^*, *\}$ — коммутативная группа.

Г3 (дистрибутивность). Для любых трех элементов a, b и c из F справедливы тождества

$$\begin{aligned}a(b + c) &= ab + ac, \\(a + b)c &= ac + bc.\end{aligned}$$

Множество R действительных чисел и множество C комплексных чисел являются полями относительно операций обычного сложения и умножения.

Поле, состоящее из конечного числа элементов, называется конечным полем, а число его элементов p — порядком поля.

Любое поле имеет следующие свойства.

Для любого элемента поля a выполняется тождество

$$a * 0 = 0 * a = 0.$$

Для любых двух ненулевых элементов поля a и b выполняется тождество

$$a * b \neq 0.$$

Для любых двух ненулевых элементов поля a и b выполняется тождество

$$-(a * b) = (-a) * b = a * (-b).$$

Если $a * b = a * c$ и $a \neq 0$, то $b = c$.

По определению поле должно содержать не меньше двух элементов.

Ниже показаны таблицы операции в поле из двух элементов (табл. 4.5).

Таблица 4.5. Операции в поле из двух элементов

+	0	1
0	0	1
1	1	0

*	0	1
0	0	0
1	0	1

Если порядок поля задан, то степень свободы при построении поля не так велика. Но если порядок поля фиксирован, конечные поля определяются одинаково с точностью до изоморфизма (отображения одного в другое).

Существует очень простой способ построения некоторых конечных полей. Если порядок поля p — простое число, то операции в поле выполняются по модулю

этого простого числа $p \pmod{p}$. Ниже показаны примеры операций в поле из трех, пяти и семи элементов (табл. 4.6–4.8).

Таблица 4.6. Пример операции в поле из трех элементов

+	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

*	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

Таблица 4.7. Пример операции в поле из пяти элементов

+	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

*	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Таблица 4.8. Пример операции в поле из семи элементов

+	0	1	2	3	4	5	6
0	0	1	2	3	4	5	6
1	1	2	3	4	5	6	0
2	2	3	4	5	6	0	1
3	3	4	5	6	0	1	2
4	4	5	6	0	1	2	3
5	5	6	0	1	2	3	4
6	6	0	1	2	3	4	5

*	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6
2	0	2	4	6	1	3	5
3	0	3	6	5	1	4	2
4	0	4	1	5	4	6	3
5	0	5	3	1	6	4	2
6	0	6	5	4	3	2	1

Легко заметить по таблицам операций, что выполняются все аксиомы и свойства полей.

Так, в поле из трех элементов «1» и «2» обратны друг другу по сложению, а «2» обратна «2» по умножению.

В поле из пяти элементов по сложению обратны друг другу «1» и «4», а также «2» и «3», а по умножению обратны друг другу «2» и «3», а также «4» и «4».

В поле из семи элементов по сложению обратны друг другу «1» и «6», «2» и «5», «3» и «4», а по умножению — «2» и «4», «5» и «3», «6» и «6».

Если подмножество F_0 поля F само является полем, то F_0 называется подполем F , а F называется расширением поля F_0 .

Можно доказать, что любое конечное поле является расширением поля из простого числа элементов p .

В любом конечном поле F порядок p единичного по умножению элемента 1 как элемента аддитивной группы поля F определяется однозначно и называется

характеристикой поля F . В поле F характеристики p для любого элемента a поля F p -кратная сумма элемента a всегда равна 0.

В дальнейшем любое конечное поле из q элементов будем называть полем Галуа и обозначать $GF(q)$. По аналогии с полем из простого числа элементов можно строить поля с числом элементов, равным любой целой степени простого числа. Только операции в таком поле выполняются над многочленами по модулю некоторого многочлена, а элемент поля задается коэффициентами многочлена. Ниже для простоты будем рассматривать поля с числом элементов, равным любой целой степени двойки.

Поле $GF(2^a)$, являющееся расширением поля $GF(2)$, содержит 2^a элементов. Все $2^a - 1$ ненулевых элементов образуют циклическую мультипликативную группу. Это значит, что в поле $GF(2^a)$ существует хотя бы один элемент b , такой, что любой ненулевой элемент $GF(2^a)$ представляет собой некоторую степень элемента b . Элемент b , обладающий этим свойством, называется примитивным элементом поля $GF(2^a)$.

Каждый элемент поля $GF(2^a)$ можно представить в виде слова длины a над полем $GF(2)$ или многочлена $f(x)$ над полем $GF(2)$, т.е. с коэффициентами из $GF(2)$, степень которого меньше a .

В этом случае сложение элементов p, q из поля $GF(2^a)$, т.е. $p + q = r$, выполняется по правилу представляющих их многочленов, $p(x) + q(x) = r(x)$ — у соответствующих многочленов складываются поразрядно коэффициенты при x , причем сложение происходит в поле $GF(2)$.

Умножение элементов в поле $GF(2^a)$, т.е. $pq = r$, выполняется по правилу представляющих эти элементы многочленов по модулю некоторого заданного многочлена $F(x)$ степени a , т.е. $p(x)q(x) = r(x) \bmod F(x)$.

Деление одного элемента p поля $GF(2^a)$ на другой элемент q поля $GF(2^a)$ соответствует умножению многочлена $p(x)$ на многочлен $r(x)$, соответствующий элементу r , обратному q , где многочлен $r(x)$ должен удовлетворять условию

$$p(x)q(x) = r(x) \bmod F(x).$$

В качестве многочлена $F(x)$ следует выбирать такой многочлен, один из корней которого является примитивным элементом поля $GF(2^a)$.

Этот примитивный элемент b представляется многочленом x или словом вида $(00 \dots 10)$, т.е.

$$b = x = (00 \dots 10).$$

Единичный элемент поля $GF(2^a)$ — b^0 записывается в виде $b^0 = 1 = (00 \dots 01)$, а все остальные элементы выражаются как степени примитивного элемента b , причем $b(2^a - 1) = b^0 = 1$.

Таким образом, при последовательном вычислении

$$b^2 = bb; b^3 = b^2b; \dots; b^{m+1} = b^m b; \dots$$

как произведения по модулю многочлена $F(x)$, определяются все элементы поля. Отсюда ясно, что поле $GF(2^a)$ полностью определяется многочленом $F(x)$. С подробным обоснованием структуры и правил построения конечных полей Галуа можно ознакомиться в [6, 41, 53, 59, 77]. Мы же ограничимся тремя примерами построения полей $GF(4)$, $GF(8)$ и $GF(16)$ (табл. 4.9–4.11).

Таблица 4.9. Поле $GF(4)$, многочлен $F(x) = x^2 + x + 1$

Номер элемента поля	Степень примитивного элемента	Соответствующий элементу многочлен	Двоичное представление элемента
1	—	0	0 0
2	b^0	1	0 1
3	b	x	1 0
4	b^2	$x + 1$	1 1

Легко убедиться, что приведенных таблиц вполне достаточно, чтобы выполнять все операции в заданном поле. Если надо складывать элементы, удобнее пользоваться их многочленным представлением, если элементы надо перемножать, удобнее пользоваться представлением через степень примитивного элемента. Если надо поделить один элемент на другой, надо умножить делимое на элемент, обратный делителю.

Таблица 4.10. Поле $GF(8)$, многочлен $F(x) = x^3 + x + 1$

Номер элемента поля	Степень примитивного элемента	Соответствующий элементу многочлен	Двоичное представление элемента
1	—	0	0 0 0
2	b^0	1	0 0 1
3	b	x	0 1 0
4	b^2	x^2	1 0 0
5	b^3	$x + 1$	0 1 1
6	b^4	$x^2 + x$	1 1 0
7	b^5	$x^2 + x + 1$	1 1 1
8	b^6	$x^2 + 1$	1 0 1

Пример для поля $GF(16)$. Пусть мы хотим сложить 10-й и 11-й элементы. Для этого мы поразрядно складываем их двоичные представления и получаем вектор (1 1 1 1), т.е. 14-й элемент. Если мы хотим эти элементы перемножить, мы пользуемся их степенным представлением и получаем $b^8 b^9 = b^{17} = b^2 b^{15} = b^2 1 = b^2$, т.е. 4-й элемент. Если мы хотим поделить 10-й элемент на 11-й, мы находим элемент, обратный 11-му по умножению, b^6 (8-й элемент), потому что $b^6 b^9 = b^{15} = 1$. Затем мы умножаем 10-й элемент на 8-й и получаем $b^8 b^6 = b^{14}$, т.е. 16-й элемент.

Отметим еще одно интересное свойство поля. Например, рассмотрим поле $GF(8)$. В таблице двоичных представлений элементов рассмотрим поведение любого разряда, например младшего, для всех элементов, кроме нулевого. Получаем вектор (1 0 0 1 0 1 1). Это так называемая последовательность максимальной длины или m -последовательность. В общем случае последовательность длины $2^a - 1$ генерируется полиномом степени a , что мы и видели при построении поля. Любые ее циклические сдвиги отличаются друг от друга в 2^{a-1} позициях. В данном

примере, если рассмотреть все возможные семь циклических сдвигов этой последовательности и добавить нулевое слово, мы получим корректирующий код из восьми слов длины 7, любые два слова которого отличаются в четырех позициях.

Таблица 4.11. Поле $GF(16)$, многочлен $F(x) = x^4 + x + 1$

Номер элемента поля	Степень примитивного элемента	Соответствующий элементу многочлен	Двоичное представление элемента
1	—	0	0 0 0 0
2	$b^0 x^4$	1	0 0 0 1
3	b	x	0 0 1 0
4	b^2	x^2	0 1 0 0
5	b^3	x^3	1 0 0 0
6	b^4	$x + 1$	0 0 1 1
7	b^5	$x^2 + x$	0 1 1 0
8	b^6	$x^3 + x^2$	1 1 0 0
9	b^7	$x^3 + x + 1$	1 0 1 1
10	b^8	$x^2 + 1$	0 1 0 1
11	b^9	$x^3 + x$	1 0 1 0
12	b^{10}	$x^2 + x + 1$	0 1 1 1
13	b^{11}	$x^3 + x^2 + x$	1 1 1 0
14	b^{12}	$x^3 + x^2 + x + 1$	1 1 1 1
15	b^{13}	$x^3 + x^2 + 1$	1 1 0 1
16	b^{14}	$x^3 + 1$	1 0 0 1

4.3.4 Введение в пространства Хемминга

Для дальнейшего рассмотрения нам понадобятся некоторые определения. Двоичным пространством Хемминга размерности n называется множество 2^n всех возможных двоичных векторов длины n . Каждый такой вектор ассоциируется с точкой в этом пространстве $\bar{x}_n = (x_1, x_2, \dots, x_n)$, $x_i \in \{0, 1\}$. По аналогии с обычным Евклидовым пространством существует n ортогональных измерений (осей), по каждому измерению (оси) возможны только два значения 0 и 1. Набор всех возможных точек пространства Хемминга называется n -мерным кубом. Естественно, возможно обобщение на q -й случай, когда все элементы принимают значения из поля $GF(q)$, но мы в дальнейшем будем рассматривать двоичный случай для простоты и экономии объема.

Весом вектора (или точки) $\bar{x}$ будем называть число ненулевых (для двоичного случая единичных) компонентов вектора и обозначать $\omega(\bar{x})$. Соответственно, расстоянием Хемминга между двумя векторами $\bar{x}$ и $\bar{y}$ $d(\bar{x}, \bar{y})$ будем называть вес вектора $\bar{z}$, являющегося результатом поразрядного сложения по модулю двух векторов $\bar{x}$ и $\bar{y}$, т.е. $d(\bar{x}, \bar{y}) = \omega(\bar{z} = \bar{x} \oplus \bar{y})$.

Сферой $Sp(r)$ радиуса r вокруг некоторой точки $\bar{x}$ будем называть множество таких точек $\bar{y}$, расстояние Хемминга которых до точки $\bar{x}$ равно r . Соответственно,

шаром $Sh(r)$ радиуса r вокруг некоторой точки $\bar{x}$ будем называть множество таких точек $\bar{y}$, расстояние Хемминга которых до точки $\bar{x}$ равно или меньше r . Площадью сферы $S_{Sp(r)}$ будем называть число точек, лежащих на этой сфере. Объемом шара $V_{Sh(r)}$ будем называть число точек, содержащихся в этом шаре. Легко подсчитать, что $S_{Sp(r)} = C_n^r$, а

$$V_{Sh(r)} = \sum_{i=0}^r C_n^i.$$

В общем случае корректирующим блочным кодом A в двоичном пространстве Хемминга размерности n будем называть набор M точек в этом пространстве, таких, что вокруг каждой точки можно провести сферы радиуса r , не касающиеся друг друга. Легко убедиться, что тогда минимальное расстояние по Хеммингу между кодовыми словами будет удовлетворять оценке $d_{\min} = 2r + 1$. Для простоты рассмотрения будем считать $k = \log_2 M$. Тогда результирующий корректирующий код будем обозначать $(n, k, d_{\min})$.

Пусть все V кодовых слов перенумерованы двоичными векторами из k бит. Эти двоичные векторы будем называть информационными векторами. Процесс кодирования состоит в замене информационного вектора длины k на кодовый вектор длины n . Соответственно, величина $R = k/n$ называется относительной скоростью кода. Далее кодовое слово $\bar{x}$ передается по двоичному симметричному каналу и на выходе канала имеем слово $\bar{y} = \bar{x} \oplus \bar{e}$, где $\bar{e}$ есть вектор ошибок, происшедших в канале.

Пусть декодер устроен следующим образом. Он вычисляет расстояние Хемминга между принятым словом $\bar{y}$ и всеми кодовыми словами $\bar{x} \in A$. В качестве декодированного выбирается то слово $\bar{x}'$, которое имеет наименьшее расстояние по Хеммингу до принятого слова. Соответствующий этому слову информационный вектор выдается получателю. Такое декодирование будем называть по минимуму расстояния.

Можно ограничить принятие решения по декодированию дополнительным сравнением результирующего слова с минимальным расстоянием r . Если оно меньше либо равно r , слово выдается получателю; если больше r , происходит отказ от декодирования — обнаруживается ошибка. Понятно, что такой алгоритм декодирования и такой код гарантированно исправляют r ошибок, происшедших в канале. Если при передаче слово остается внутри шара, соответствующего переданному слову, декодирование происходит правильно. Если принятое слово попадает в другой шар (соответствующий другому кодовому слову) — происходит ошибочное декодирование. Если слово не попадает ни в какой шар, происходит обнаружение ошибки.

Можно искусственно уменьшить радиус сферы, по которому принимается решение о декодировании. Пусть решение принимается по величине $r' < r$. Тогда мы будем говорить, что код и данный алгоритм декодирования гарантированно исправляют r' ошибок и дополнительно обнаруживают $2r - r'$ ошибок, причем $d_{\min} = r' + (2r - r') + 1 = 2r + 1$. Другой крайней границей (см. примеры в начале главы) является отсутствие исправления и обнаружение $2r$ ошибок.

В рассмотренном в начале главы примере в коде-повторении рассматриваются две сферы — вокруг нулевого и единичного слова, а в коде с проверкой на чет-

ность кодовыми сферами являются половина всех точек пространства Хемминга с четным весом.

Отметим также, что существуют коды, для которых сферы радиуса r занимают все пространство Хемминга — точек между сферами не остается. Это свойство выполняется при выполнении следующего условия (равенство объемов 2^k шаров радиуса r и объема всего пространства Хемминга)

$$2^k \sum_{i=0}^r C_n^i = 2^n.$$

Это условие называется границей Хемминга, а коды, соответствующими этому условию, — лежащими на границе Хемминга, или совершенными кодами. В случае декодирования совершенных кодов по расстоянию r вероятность обнаружения ошибки будет равна 0 — будет либо ошибочное, либо правильное декодирование.

4.3.5 Линейные коды

Линейные коды составляют лишь небольшой подкласс блочных кодов, однако благодаря своей красивой структуре они стали основным объектом в теории кодирования. Пусть V_n — векторное пространство размерности n над полем $GF(q)$. Подпространство размерности k пространства V_n называется q -м линейным кодом длины n с k информационными символами или (n, k) кодом.

Если A — линейный код, то подпространство A будем называть линейным подкодом A . При $q = 2$ линейные коды называются групповыми. Пусть $d_{\min}$ — минимальное расстояние линейного кода A . Тогда для любого ненулевого вектора $\bar{v} \in A$ выполняется неравенство $\omega(\bar{v}) = \omega(\bar{v} - \bar{0}) \geq d_{\min}$, так как нулевой вектор также принадлежит коду A . Для определения линейного кода удобно пользоваться матрицей G . Пусть $\bar{g}_1, \bar{g}_2, \dots, \bar{g}_k$ есть некоторый базис (n, k) кода A и пусть матрица G — матрица из k строк и n столбцов, i -й строкой которой является базисный вектор $\bar{g}_i$. Матрица G называется порождающей матрицей кода A . Из свойства базиса следует, что любой кодовый вектор может быть представлен как линейная комбинация строк $\bar{g}_1, \bar{g}_2, \dots, \bar{g}_k$ матрицы G и, наоборот, любая линейная комбинация строк $\bar{g}_1, \bar{g}_2, \dots, \bar{g}_k$ матрицы G представляет собой кодовый вектор и, более того, различные кодовые комбинации представляют собой различные кодовые векторы. Общее число кодовых векторов равно q^k .

Для задания линейных кодов также используется матрица H , которая определяется следующим образом. Пусть A_0 — множество векторов $\bar{y} = (y_1, y_2, \dots, y_n)$, таких, что для любого $\bar{x} = (x_1, x_2, \dots, x_n) \in A$ выполняется равенство

$$\bar{y}\bar{x}^T = \bar{0} \text{ (т. е. } \sum_{i=1}^n y_i x_i),$$

где индекс T означает транспонирование.

Как известно, A_0 является подпространством размерности $(n - k)$ векторного пространства размерности n . Это подпространство называется подпространством, двойственным коду A . Порождающую матрицу H двойственного кода A_0 назовем проверочной матрицей кода A . Проверочная матрица имеет размер

$(n - k) \times n$ и, по определению, есть

$$H\bar{x}^T = \bar{0} \Leftrightarrow \bar{x} \in A.$$

В частности, если взять в качестве векторов $\bar{x}$ базисные векторы кода A , то получим

$$HG^T = \bar{0}.$$

Поскольку по определению ранг матрицы H равен $(n - k)$, то H содержит невырожденную подматрицу H' размера $(n - k) \times (n - k)$. Путем перестановки столбцов H и соответствующего изменения нумерации компонент кодовых векторов всегда можно добиться того, чтобы матрица H' была образована последними $(n - k)$ столбцами H . С помощью элементарных операций над строками матрица H всегда может быть приведена к следующему виду

$$H_0 = [P, I_{n-k}],$$

где P — некоторая матрица размера $(n - k) \times k$, а I_{n-k} — единичная матрица размера $(n - k) \times (n - k)$.

Поскольку матрица H_0 получена элементарными операциями над строками, то

$$H\bar{x}^T = \bar{0} \Leftrightarrow H_0\bar{x}^T = \bar{0},$$

т. е. матрица H_0 также называется проверочной матрицей кода A .

Обозначим через p_{ij} элемент подматрицы P матрицы H_0 , стоящий в i -й строке и j -м столбце. Очевидно, что необходимым и достаточным условием принадлежности вектора $\bar{x} = (x_1, x_2, \dots, x_n)$ коду A является выполнение следующей системы неравенств

$$x_{r+i} = - \sum_{j=1}^k p_{ij} x_j, \quad 1 \leq i \leq n - k.$$

Таким образом, если компоненты $x_1, x_2, \dots, x_k \in GF(q)$ заданы, то существует ровно один кодовый вектор, первыми k компонентами которого являются $x_1, x_2, \dots, x_k$, и можно вычислить по проверочным соотношениям остальные компоненты этого вектора. Первые k символов называются информационными, а остальные $(n - k)$ — проверочными.

Справедливо следующее утверждение. Пусть проверочная матрица H_0 линейного (n, k) кода A имеет вид $H_0 = [P, I_{n-k}]$. Тогда матрица $G_0 = [I_r - P^T]$ является порождающей матрицей кода A . Справедливо также и обратное утверждение.

Можно доказать следующее утверждение. Минимальный вес линейного (n, k) кода A равен d тогда и только тогда, когда любые $(d - 1)$ столбцов проверочной матрицы этого кода линейно независимы, но некоторые d столбцов этой матрицы линейно зависимы.

Следствием этого является неравенство $d \leq n - k + 1$.

Пусть $\bar{x}$ — переданный кодовый вектор по двоичному симметричному каналу, $\tilde{\bar{x}} = \bar{x} \oplus \bar{e}$ — принятый вектор, а $\bar{e}$ — вектор ошибок. Тогда вектор $\bar{s} = \tilde{\bar{x}} H^T$ называется синдромом. По определению проверочной матрицы

$$\bar{s} = \tilde{\bar{x}} H^T = (\bar{x} \oplus \bar{e}) H^T = \bar{x} H^T \oplus \bar{e} H^T = \bar{e} H^T.$$

Векторы $\bar{x}, \bar{y}$ называются сравнимыми по модулю A , если $(\bar{x} - \bar{y}) \in A$; сравнимость векторов определяется следующим образом:

$$\bar{x} \equiv \bar{y} \pmod{A}.$$

Из определения сравнимости векторов следует, что:

$$\begin{aligned} \bar{x} &\equiv \bar{x} \pmod{A}; \\ \text{если } \bar{x} &\equiv \bar{y} \pmod{A}, \text{ то } \bar{y} \equiv \bar{x} \pmod{A}; \\ \text{если } \bar{x} &\equiv \bar{y}, \bar{y} \equiv \bar{z} \pmod{A}, \text{ то } \bar{x} \equiv \bar{z} \pmod{A}. \end{aligned}$$

В этом случае все пространство Хемминга размерности n можно разбить на непесекающиеся классы таким образом, что любые два вектора из одного класса будут сравнимы по модулю A . Эти классы называются смежными классами. Смежный класс, содержащий вектор $\bar{x}$, обозначается через $(\bar{x})$. Смежный класс $(\bar{0})$ совпадает с кодом A . Каждый смежный класс состоит из q^k векторов, а всего смежных классов q^{n-k} . Для всех векторов из одного смежного класса синдром $\bar{s}$ одинаков. Вектор смежного класса с минимальным весом называется образующим (лидером) смежного класса.

Легко заметить, что для представления порождающей матрицы в таком виде кодирование сводится к умножению информационного слова на порождающую матрицу кода. Также важно заметить, что в случае линейных кодов минимальный ненулевой вес кодового слова совпадает с минимальным кодовым расстоянием.

Для декодирования линейного кода можно построить таблицу взаимнооднозначного соответствия лидеров смежного класса и соответствующих им синдромов. Тогда алгоритм декодирования имеет следующий вид.

1. По принятому вектору вычисляется синдром.
2. По синдрому определяется лидер смежного класса.
3. В случае двоичного кода лидер смежного класса поразрядно суммируется по модулю 2 с принятым словом и результат выдается получателю.

Этот алгоритм декодирования полностью соответствует описанному выше алгоритму по расстоянию. Для понимания этого факта заметим, что если мы имеем таблицу смежных классов, где в качестве первой строки записан сам код A , а дальнейшие лидеры смежных классов расположены по росту веса Хемминга, то каждому столбцу соответствуют векторы из шара радиуса r вокруг слова из первой строки вместе с некоторыми векторами вне шара (для несовершенных кодов). Таблица записи смежных классов называется таблицей стандартного расположения кода. В ней, как указывалось, первые $V_{Sh(r)} = \sum_{i=0}^r C_n^i$ строк соответствуют исправлению ошибок, а остальные $q^{n-k} - V_{Sh(r)} = \sum_{i=0}^r C_n^i$ строк — обнаружению.

Интересно отметить одно важное преимущество линейных кодов. Теперь нет необходимости запоминать кодовую таблицу при кодировании. Для обнаружения ошибок в принятом слове также достаточно умножения на матрицу.

Рассмотрим следующий пример линейного кода. Пусть проверочная матрица H представляется в следующем виде

$$H = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

Путем линейных комбинаций приведем ее к систематическому виду (с правой единичной матрицей)

$$H_0 = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

Из проверочной матрицы в систематическом виде H_0 получим порождающую матрицу кода

$$G_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Получился линейный код (7,4), причем, поскольку любые два столбца проверочной матрицы линейно независимы (разные), минимальное расстояние кода по Хеммингу $d_{\min} = 3$ и код имеет параметры (7,4,3). Это так называемый код Хемминга и совершенный код. Код имеет 16 кодовых слов. Ниже в табл. 4.12 показано разбиение пространства Хемминга размерности 7 на смежные классы, где первая строка таблицы есть сам код. Всего имеется 8 смежных классов. Также в таблице слева показан соответствующий синдром.

Таблица 4.12. Разбиение пространства Хемминга размерности 7 на смежные классы

Синдром кода	1-е слово	2-е слово	3-е слово	4-е слово	5-е слово	6-е слово
000	0000000	1000011	0100101	1100110	0010110	1010101
011	1000000	0000011	1100101	0100110	1010110	0010101
101	0100000	1100011	0000101	1000110	0110110	1110101
110	0010000	1010011	0110101	1110110	0000110	1000101
111	0001000	1001011	0101101	1101110	0011110	1011101
100	0000100	1000111	0100001	1100010	0010010	1010001
010	0000010	1000001	0100111	1100100	0010100	1010111
001	0000001	1000010	0100100	1100111	0010111	1010100

Видно, что код является совершенным. В каждом столбце под соответствующим кодовым словом содержатся все (семь) слова сферы радиуса 1 с центром в этом слове. Синдром кода соответствует тому столбцу проверочной матрицы, на позиции которого произошла ошибка.

Также видно еще одно очень важное свойство линейных кодов — спектр расстояний от нулевого слова до всех других кодовых слов такой же, как и от любого другого кодового слова. Тогда очень легко подсчитать вероятность ошибки декодирования при условии, что передается нулевое слово и исправляется одиночная ошибка. Ошибочное декодирование произойдет тогда, когда принятое слово попадет в сферу любого другого кодового слова. Для этого достаточно подсчитать вес всех слов во всех шарах кроме первого (правильного).

Продолжение табл. 4.12

Синдром кода	7-е слово	8-е слово	9-е слово	10-е слово	11-е слово	12-е слово
000	0110011	1110000	0001111	1001100	0101010	1101001
011	1110011	0110000	1001111	0001100	1101010	0101001
101	0010011	1010000	0101111	1101100	0001010	1001001
110	0100011	1100000	0011111	1011100	0111010	1111001
111	0111011	1111000	0001111	1000100	0100010	1100001
100	0110111	1110100	0001011	1001000	0101110	1101101
010	0110001	1110010	0001101	1001110	0101000	1101010
001	0110010	1110001	0001110	1001101	0101011	1101000

Окончание табл. 4.12

Синдром кода	13-е слово	14-е слово	15-е слово	16-е слово
000	0011001	1011010	0111100	1111111
011	1011001	0011010	1111100	0111111
101	0111001	1111010	0011100	1011111
110	0001001	1001010	0101100	1101111
111	0010001	1010010	0110100	1110111
100	0011101	1011110	0111000	1111011
010	0011011	1011000	0111110	1111101
001	0011000	1011011	0111101	1111110

Для этого кода и алгоритма декодирования вероятность обнаружения ошибки $P_B = 0$, а вероятность ошибки декодирования вычисляется по формуле

$$P_B = 21p^2(1-p)^5 + 30p^3(1-p)^4 + 35p^4(1-p)^3 + 26p^5(1-p)^2 + 7p^6(1-p) + p^7.$$

Подчеркнем, что это есть точное вычисление вероятности ошибки декодирования в соответствии с таблицей стандартного расположения, т. е. спектром весов кода.

Если модифицировать алгоритм, как в примере в начале главы, и уменьшить радиусы сфер от единицы до нуля, тогда вероятность ошибки декодирования будет равна

$$P_B = 7p^3(1-p)^4 + 7p^4(1-p)^3 + p^7.$$

Вероятность обнаружения ошибки в этом случае равна

$$P_B = 21p^2(1-p)^5 + 23p^3(1-p)^4 + 28p^4(1-p)^3 + 26p^5(1-p)^2 + 7p^6(1-p).$$

В общем случае двоичные коды Хемминга имеют параметры $2^m - 1$, $2^m - 1 - m$, 3 для любого целого m . Проверочная матрица любого кода Хемминга состоит из всех различных ненулевых столбцов длины m . В пространстве Хемминга шары радиуса 1 вокруг всех кодовых слов заполняют все пространство Хемминга, т. е. код Хемминга является совершенным и для него выполняется граница Хемминга. В случае кодов Хемминга граница Хемминга принимает следующий вид:

$$2^{2^m-1-m} \sum_{i=0}^1 C_{2^m-1}^i = 2^{2^m-1} \Rightarrow \sum_{i=0}^1 C_{2^m-1}^i = 2^m \Rightarrow 1 + 2^m - 1 = 2^m.$$

4.3.6 Циклические коды

Циклические коды являются подклассом линейных кодов, которые не изменяют своих свойств при циклической перестановке компонент кодовых слов. Благодаря этому свойству процедура кодирования и простейших процедур декодирования по сложности линейная, а не квадратичная, как в случае линейных кодов. Все совершенные коды, рассмотренные ниже конкретные классы кодов — Рида-Маллера, Боуза-Чоудхури-Хоквингема, Рида-Соломона — имеют циклические аналоги.

Пусть A представляет собой q -й циклический (n, k) код. Если вместе с каждым кодовым вектором $\bar{a} = (a_0, a_1, \dots, a_{n-1})$ коду также принадлежит вектор $\bar{a}_{\text{Ц}} = (a_{n-1}, a_0, a_1, \dots, a_{n-2})$, то код называется циклическим. Кодовый вектор удобно также представлять в качестве многочлена $a(x) = a_0 + a_1x + a_2x^2 + \dots + a_{n-1}x^{n-1}$ над полем $GF(q)$. Тогда вектору $\bar{a}_{\text{Ц}} = (a_{n-1}, a_0, a_1, \dots, a_{n-2})$ соответствует многочлен $xa(x) - a_{n-1}(x^n - 1)$. Первая и последняя компоненты циклического кода рассматриваются как соседние, поэтому степени x^0 и x^n тождественны.

Различные многочлены $a(x)$ и $b(x)$ степени $n - 1$ и менее принадлежат различным классам вычетов по модулю $x^n - 1$. Класс вычетов по модулю $x^n - 1$, содержащий многочлен $a(x)$, будем называть $(a(x))$. Циклическому сдвигу $\bar{a}_{\text{Ц}}$ соответствует класс вычетов $(xa(x)) = (x)(a(x))$. Множество классов вычетов с соответствующими операциями по модулю многочлена $x^n - 1$ будем обозначать R_n . Если при представлении кодовых векторов в виде последовательностей необходимо рассматривать циклические перестановки (соответствующие операции над матрицами), то при операциях над R_n достаточно лишь линейных по сложности операций над многочленами.

Пусть I — подмножество R_n , соответствующее коду A , а именно:

$$I = \{a(x) \in R_n \Leftrightarrow \bar{a} \in A\}.$$

Тогда для произвольного многочлена $(b(x)) \in R_n$ выполняется тождество $(b(x)a(x)) \in I$, т. е. I является идеалом в R_n .

Пусть $g(x)$ является ненулевым нормированным по старшей степени многочленом минимальной степени, таким, что $(g(x)) \in I$. Пусть $b(x)$ — произвольный многочлен, принадлежащий идеалу I . Тогда, если $r(x)$ — остаток от деления $b(x)$ на $g(x)$, то

$$b(x) = a(x)g(x) + r(x)$$

и

$$r(x) = (b(x)) - (a(x))(g(x)).$$

Отсюда следует, что $r(x) \in I$.

Но тогда многочлен $r(x)$ должен быть нулевым, т. е. $b(x)$ делиться на $g(x)$.

Таким образом, практически доказана следующая теорема.

Пусть A — циклический (n, k) код и $I = \{a(x) \in R_n \Leftrightarrow \bar{a} \in A\}$. Тогда существует такой нормированный многочлен $g(x)$ степени $n - k$, делящий $x^n - 1$, что выполнение сравнения $b(x) \equiv 0 \pmod{g(x)}$ является необходимым и достаточным условием того, что $b(x) \in I$. Многочлен $g(x)$ определяется однозначно и называется порождающим многочленом циклического кода A .

Пусть $g(x) = g_0 + g_1x + \dots + g_{n-k}x^{n-k}$ — порождающий многочлен циклического кода A . Пусть G — матрица из k строк и n столбцов следующего вида:

$$G = \begin{bmatrix} g_0 & g_1 & \cdot & \cdot & \cdot & g_{n-k} & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & g_0 & g_1 & \cdot & \cdot & \cdot & g_{n-k} & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & g_0 & g_1 & \cdot & \cdot & \cdot & g_{n-k} & 0 \\ 0 & \cdot & \cdot & \cdot & \cdot & g_0 & g_1 & \cdot & \cdot & \cdot & g_{n-k} \end{bmatrix}.$$

Строки матрицы G являются кодовыми словами, а ранг равен k . Следовательно, это порождающая матрица кода A . Порождающий многочлен $g(x)$ является кратким представлением матрицы. Многочлен $h(x) = (x^n - 1)/g(x)$ называется проверочным многочленом кода. Этот многочлен порождает код A_0 , двойственный к коду A .

Пусть $h(x) = h_0 + h_1x + \dots + h_kx^k$. Справедлива следующая теорема.

Двойственный код A_0 циклического (n, k) кода A также является циклическим. Порождающим многочленом кода A_0 является многочлен вида $h_0^{-1}x^k h(x^{-1})$, где $h(x)$ — проверочный многочлен кода A .

Из теоремы следует, что следующая матрица является порождающей матрицей кода A и проверочной матрицей кода A_0 :

$$H = \begin{bmatrix} h_k & h_{k-1} & \cdot & \cdot & \cdot & h_0 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & h_k & h_{k-1} & \cdot & \cdot & \cdot & h_0 & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & h_k & h_{k-1} & \cdot & \cdot & \cdot & h_1 & 0 \\ 0 & \cdot & \cdot & \cdot & 0 & h_k & h_{k-1} & \cdot & \cdot & \cdot & h_1 \end{bmatrix}.$$

Многочлен $b(x) = x^{n-k}b_0(x) + r(x)$, где $b_0(x)$ — информационный многочлен, является кодовым тогда и только тогда, когда он делится на порождающий многочлен $g(x)$. Понятно, что в качестве многочлена проверочных символов можно взять остаток от деления $x^{n-k}b_0(x)$ на $g(x)$, заменив все коэффициенты на противоположные. Это фактически задает алгоритм кодирования по порождающей матрице кода (порождающему многочлену).

Аналогичным образом можно кодировать по проверочной матрице кода (проверочному многочлену). Естественно, аналогичным образом можно вычислять

синдром кода. Более подробно теория циклических кодов изложена в классических монографиях [6, 41, 53, 59, 77].

Существуют элементарные операции над кодами. К числу самых важных из них относятся укорочение кода и добавление общей проверки на четность. Укорочение кода за счет выкалывания первых m информационных координат сводит (n, k) код с расстоянием $d_{\min}$ к $(n - m, k - m)$ коду с расстоянием не меньше $d_{\min}$. Это бывает очень важно при согласовании источника сообщений с конкретными параметрами кода. Добавление общей проверки на четность сводит (n, k) код с расстоянием $d_{\min} = 2r + 1$ к $(n + 1, k)$ коду с расстоянием на единицу большим.

4.3.7 Наиболее известные классы блочных кодов

Коды Боуза–Чоудхури–Хоквингема

Коды Боуза–Чоудхури–Хоквингема (БЧХ) составляют один из больших классов кодов, исправляющих независимые ошибки веса t , метод построения которых может быть явно задан. Пусть m — произвольное целое положительное число и n — один из делителей $q^m - 1$. В поле $GF(q^m)$ всегда существует элемент порядка n . Пусть b — один из таких элементов. Заметим, что все элементы b^i ($0 \leq i < n$) различны. Можно доказать [41, 53, 59, 77] следующую теорему.

Если порождающий многочлен $g(x)$ циклического кода A длины n имеет своими корнями $b^{l+1}, b^{l+2}, \dots, b^{l+r}$ ($1 \leq r < n$), где l ($0 \leq l < n$) — некоторое целое число, то минимальное расстояние этого кода не меньше чем $r + 1$.

Расстояние $r + 1$, гарантируемое этой теоремой, называется нижней границей кодового расстояния БЧХ кодов. Чтобы код гарантированно исправлял t ошибок, надо положить $r = 2t$. Тогда минимальное расстояние кода будет удовлетворять оценке $d_{\min} \geq 2t + 1$. Эту границу называют конструктивным расстоянием кода. Реальное расстояние может быть больше. Легко убедиться, что число проверочных символов кода не превосходит величины $2mt$.

В частном случае $q = 2$, $n = 2^m - 1$, $l = 0$, $r = 2t$ БЧХ коды называются примитивными или БЧХ кодами в узком смысле. При $t = 1$ — это циклические коды Хемминга.

Рассмотрим алгоритм построения двоичных кодов БЧХ с кодовым расстоянием не менее заданного. Для нечетных $d_{\min} = 2t + 1$ элементы $b, b^2, \dots, b^{2t}$ являются корнями $g(x)$ и для четных $d_{\min} = 2t + 2$ элементы $1, b, b^2, \dots, b^{2t}$ являются корнями $g(x)$. В обоих случаях порождающий многочлен $g(x)$ есть произведение некоторых минимальных функций $m_i(x)$ [41, 53, 59, 77]. В соответствии со свойствами минимальных функций каждая четная степень элемента b , большая нуля, является корнем минимальной функции для некоторой предшествующей нечетной степени элемента b . Например, b^2, b^4 — корни $m_1(x)$, b^6 — корень $m_3(x)$, b^{10} — корень $m_5(x)$ и т. д. Следовательно, порождающий многочлен можно получить по формуле

$$g(x) = \left\{ \begin{array}{l} \text{НОК}(m_1(x)m_3(x)\dots m_{2t-1}(x)), d_{\min} = 2t + 1, \\ \text{НОК}(m_0(x)m_1(x)m_3(x)\dots m_{2t-1}(x)), d_{\min} = 2t + 2. \end{array} \right\}.$$

Таким образом, построение БЧХ кода с кодовым расстоянием не менее заданного состоит в следующем.

Задается конструктивное кодовое расстояние $d_{\min} = 2t + 1$ или $d_{\min} = 2t + 2$ и определяется соответствующее значение t .

Находятся все минимальные функции $m_{2i-1}(x)$, $i = 1, \dots, t$.

Определяется порождающий многочлен $g(x)$, степень которого равна числу проверочных символов. Число информационных символов находится из соотношения $k = n - r - 2^m - 1 - \deg g(x)$.

Проверяется, не делится ли полученный многочлен $g(x)$ на $m_{2t+1}(x)$. Если нет, построение закончено. Если делится, проверяется многочлен большей степени, и т. д. Таким образом, находится реальное расстояние $d_{\min}$.

Коды Рида–Соломона

Коды Рида–Соломона (РС) представляют собой частный случай недвоичных циклических кодов БЧХ. Эти коды, как и двоичные, задаются порождающим многочленом $g(x)$ при $l = 0$ и $m = 1$.

Таким образом, коды имеют длину $n = q - 1$ и порождающий многочлен $g(x)$ с коэффициентами из $GF(q)$ должен содержать в качестве своих корней элементы $b, b^2, b^3, \dots, b^{d_{\min}-1}$, где b — примитивный элемент $GF(q)$. Порождающий многочлен определяется следующим соотношением:

$$g(x) = \prod_{i=1}^{d_{\min}-1} (x + b^i).$$

Таким образом, число проверочных символов равно степени порождающего многочлена $d_{\min} - 1$, а число информационных символов $k = n - d_{\min} + 1$.

Как следует из определения линейных кодов, всегда имеется в коде слово веса 1 (только с одной единицей на информационных позициях). Если предположить, что на всех проверочных позициях такого слова также единицы, вес этого слова равен $n - k + 1$. Следовательно, в линейном коде минимальное расстояние не может быть больше величины $n - k + 1$. Поэтому РС коды являются кодами с максимально возможным минимальным расстоянием. Естественно, максимальность расстояния достигается за счет недвоичной структуры кода, т. е. двоичных кодов с аналогичными параметрами n и k не существует.

Декодирование кодов БЧХ и РС

Для кодов БЧХ и РС существуют алгебраические процедуры декодирования, сводящиеся в основном к следующей вычислительной процедуре [6, 7].

1. По принятому слову вычисляется $d_{\min} - 1$ значений «частичных» синдромов. Если все синдромы равны нулю, принятое слово является кодовым.
2. Вычисляется многочлен локаторов стираний.
3. Вычисляется многочлен обобщенных проверок.
4. Вводится многочлен локаторов ошибок.

5. Многочлен локаторов ошибок определяется в результате решения ключевого уравнения. Рекурсивная процедура решения этого уравнения подробно изложена в [6, 7].
6. Определяются локаторы ошибок, т.е. те места, где произошли ошибки.
7. Определяются значения ошибок.

В общем случае алгоритм позволяет исправлять t ошибок и τ стираний, причем $d_{\min} = t + \tau + 1$, где $\tau \geq t$.

Достаточно давно существуют вычислительные и аппаратные реализации декодирования (естественно, и кодирования) БЧХ и РС кодов достаточно большой длины.

Коды Рида-Маллера

Коды Рида-Маллера (РМ) существуют в обобщенном аналоге, но мы приведем только упрощенный двоичный случай. Двоичные коды РМ существуют в широкой области скоростей, минимальных расстояний и длин. Они эквивалентны циклическим кодам с общей проверкой на четность. Они являются основой для мажоритарно декодируемых кодов.

Для любых $m, r < m$, существует РМ код, для которого

$$\begin{aligned} n &= 2^m, \\ k &= 1 + C_m^1 + \dots + C_m^r, \\ n - k &= 1 + C_m^1 + \dots + C_m^{m-r+1}, \\ d_{\min} &= 2^{m-r}. \end{aligned}$$

Такой код называется РМ кодом r -го порядка.

Обозначим векторы

$$\begin{aligned} \bar{a}_0 &= (00000000 \dots 00000000), \\ \bar{a}_1 &= (01010101 \dots 01010101), \\ \bar{a}_2 &= (00110011 \dots 00110011), \\ \bar{a}_3 &= (00001111 \dots 00001111), \\ \bar{a}_4 &= (00000000 \dots 11111111) \end{aligned}$$

и т.д.

Код РМ порядка r определяется как код, базисом которого являются все векторы $\bar{a}_0, \bar{a}_1, \bar{a}_2, \dots, \bar{a}_m$ и все векторные произведения r или меньшего числа этих векторов. Легко убедиться, что коды РМ порядка r и $m - r - 1$ двойственны или дуальны друг другу. Так, например, РМ код нулевого ($r = 0$) порядка есть не что иное, как код-повторение $(n, 1, n)$. Двойственный ему код $(m - 1)$ -го порядка есть $(n, n - 1, 2)$ -код с проверкой на четность. Код первого порядка $(2^m, m + 1, 2^{m-1})$ есть код на основе m -последовательности, рассмотренный выше, с общей проверкой на четность. Ему двойственен код $(m - 2)$ -го порядка $(2^m, 2^m - m - 1, 4)$, являющийся кодом Хемминга с общей проверкой на четность. Все эти коды допускают интересные геометрические интерпретации [41, 77] и различные алгоритмы декодирования.

4.3.8 Итеративные и каскадные коды

С одной стороны, как гласит теорема Шеннона для канала с шумами (см. раздел 4.3.2), необходимо стремиться для повышения пропускной способности к большой длине кода. С другой стороны, при увеличении длины кода существенно возрастает сложность алгоритмов декодирования. Даже для кодов БЧХ она хотя и не экспоненциальная, но, тем не менее, растет от длины как некоторая степень полинома. Все это приводит к затруднительности реализации очень длинных кодов.

Для преодоления этой проблемы в самом начале развития теории кодирования Элайесом [113] были предложены так называемые произведения кодов. Начнем с примера. Пусть имеется код с проверкой на четность длины 3: (3,2,2). Словами кода являются все слова четного веса длины 3. Таких слов 4. Это слова: (000), (011), (101), (110). Минимальное расстояние кода равно двум. Будем рассматривать этот код как элемент построения кода-произведения или итеративного кода. Для этого будем поступать следующим образом. Запишем все информационные слова в виде матрицы размером 2×2 . Каждую строчку этой матрицы закодируем кодом (3,2,2). Получаем матрицу размером 2×3 . Далее, каждый столбец этой матрицы также закодируем кодом (3,2,2). Получаем матрицу размером 3×3 .

Тем самым мы получили линейный алгоритм преобразования матрицы 2×2 (информационного слова длины 4) в матрицу размером 3×3 (кодое слово длины 9), т.е. линейный (9,4) код. Такой код будем называть произведением кодов (3,2,2) или итеративным кодом, в котором первый код (строчный) (3,2,2) является внешним, а второй код (столбцовый) (3,2,2) является внутренним.

Оценим минимальный ненулевой вес кодового слова такого кода. Если кодое слово будет содержать одну единицу на информационных позициях, то это приведет к трем единицам на проверочных:

$$\begin{bmatrix} 00 & 0 \\ 10 & 1 \\ 10 & 1 \end{bmatrix}.$$

Если на информационных позициях две единицы, возможны две ситуации.

У этих двух единиц есть общая строка или столбец. Тогда вес кодового слова также равен четырем:

$$\begin{bmatrix} 00 & 0 \\ 11 & 0 \\ 11 & 0 \end{bmatrix}.$$

У этих двух единиц нет общей строки или столбца. Тогда вес кодового слова еще больше (6):

$$\begin{bmatrix} 10 & 1 \\ 01 & 1 \\ 11 & 0 \end{bmatrix}.$$

Также легко убедиться, что если на информационных позициях три единицы, то, как минимум, одна приведет к появлению проверочной единицы. Таким образом, в этом случае вес кодового слова будет также не менее четырех. Тем

самым мы показали, что вес кодового вектора, а следовательно, и минимальное кодовое расстояние такого кода будет не менее четырех, т. е. мы построили $(9,4,4)$ код. Существует лучший код — код Хемминга с общей проверкой на четность — $(8,4,4)$, однако мы с самого начала не настаивали на оптимальности кодов.

Существует очень простой алгоритм декодирования такого кода. Для этого достаточно выполнить все столбцовые и строчные проверки на четность кода. Если все проверки удовлетворяются (нулевые), то принятое слово кодовое. Оно и выдается получателю. Если по одной строчной и столбцовой проверке не удовлетворяются, то на пересечении этих проверок исправляется ошибка. Если в столбцах или строках имеется более чем по одной неудовлетворенной проверке, то обнаруживается ошибка. Легко убедиться, что такой алгоритм гарантированно исправляет одиночную ошибку и обнаруживает двойную, т. е. реализует кодовое расстояние такого кода. Очевидно, что все рассуждения остаются в силе для образующих кодов $(n, n-1, 2)$ произвольной длины. Получается результирующий $(n^2, n^2+1-2n, 4)$ код. Аналогичным образом можно строить трех- и четырехмерные коды и коды произвольной размерности. Так, в нашем примере получается $(27,8,8)$ код. Видно, что у таких кодов все кодовые параметры — длина, число информационных символов и минимальное кодовое расстояние — являются произведением соответствующих параметров образующих кодов.

В общем случае имеется внешний строчный код (n_2, k_2, d_2) и внутренний строчный код (n_1, k_1, d_1) . Их произведение является итеративным кодом

$$(n_1 n_2, k_1 k_2, d_1 d_2).$$

Самый простой алгоритм декодирования такого кода состоит в попеременном построчном и постолбцовом декодировании таких кодов, однако он не реализует кодового расстояния. Для реализации кодового расстояния нужны более сложные алгоритмы.

Следует отметить еще одно очень важное достоинство таких кодов. При кодировании символы записываются в таблицу построчно, а считываются в канал столбцами. Это приводит к так называемому перемежению, если в канале возникают пакетированные ошибки, при декодировании они эффективно исправляются.

Каскадные коды были предложены Форни [96] и являются модификацией итеративных кодов. Исходные коды обозначаются также как и в итеративных, а вот внешний является q -м, где символом внешнего кода является информационное слово внутреннего кода, т. е. $q = 2^{k_1}$. Наиболее часто в качестве внешних используются оптимальные недвойчные коды — РС коды. Алгоритм декодирования состоит в декодировании внутреннего столбцового кода и внешнего РС кода, причем, как правило, декодирование внутреннего кода согласовано с каналом (носит вероятностный характер), а внешнее декодирование является алгебраическим. Результаты оценки помехоустойчивости такой каскадной пары будут приведены в конце главы при сравнении с турбокодами.

4.3.9 Мягкое декодирование, энергетический выигрыш кодирования — основные определения

Энергетический выигрыш кодирования (ЭВК) G определяет выигрыш по помехоустойчивости при применении корректирующего кода в конкретной систе-

ме. Как правило, в системах предполагают применение символов с двухпозиционной фазовой модуляцией ФМ2, использующей противоположные сигналы (см. ниже). В отличие от пропускной способности произвольного канала, задаваемой при стремящейся к нулю вероятности ошибки, ЭВК определяется для фиксированной вероятности ошибки декодирования. В стандартной ситуации ЭВК рассматривается для фиксированной вероятности ошибки в бите, а не кодовом слове, при этом дополнительное обнаружение ошибки также не учитывается. Для получения заданного значения вероятности ошибочного приема одного символа P_b в информационной последовательности надо обеспечить на выходе демодулятора приемника некоторое необходимое минимально допустимое отношение сигнал/шум. При передаче информации с корректирующим кодированием уже вместо k информативных символов за заданное время требуется передача n символов с добавлением проверочных за то же время при том же уровне мощности сигналов. При этом придется сокращать длительность символов при передаче в n/k раз, что потребует расширения полосы частот в n/k раз. Исходное заданное значение вероятности ошибки в канале p будет обеспечиваться уже при другом отношении сигнал/шум. Разница отношений сигнал/шум при применении корректирующего кода и без него при ее положительном значении определяет ЭВК, выражаемый в децибелах. Обычно ЭВК растет с уменьшением требуемой вероятности ошибки P_b , причем существует практически в любой системе граничное отношение сигнал/шум, начиная с которого корректирующий код дает выигрыш, т. е. ЭВК положителен.

Быстрая ориентировочная оценка энергетической эффективности кодирования для целей оперативного сравнения кодов производится по асимптотическому энергетическому выигрышу от кодирования (АЭВК) $G_A = 10 \lg R d_{\min}$ (дБ), где $R = k/n$ — относительная скорость кода; $d_{\min}$ — минимальное кодовое расстояние. Величина АЭВК G_A характеризует ЭВК при вероятности $P_b \rightarrow 0$ и является верхней границей реального ЭВК. Выигрыш от кодирования может быть использован наиболее эффективным способом, например путем уменьшения мощности передатчиков в системах связи, уменьшения размеров антенн или увеличения скорости передачи.

Следует отметить, что такое определение ЭВК не учитывает частотной эффективности системы — наиболее эффективной может оказаться система с очень большим расширением полосы сигнала. Кроме этого, использование двоичной модуляции и корректирующего кода с относительной скоростью R приводит к существенному ограничению пропускной способности канала связи. Так, на рис. 4.4 показано сравнение пропускной способности канала в случае произвольного канала с аддитивным гауссовым шумом (Shannon Bound) и канала с двоичной модуляцией и корректирующим кодом с относительной скоростью R (BPSK)

На рис. 4.4 в случае двоичной модуляции рассматривается так называемый полунепрерывный канал, в котором в отличие от канала с «жестким» решением не принимаются решения о принятом символе. Точнее говоря, решение о принятом символе может и приниматься, но декодеру сообщается вся информация, содержащаяся в сигнале, соответствующем принятому символу. Такой алгоритм декодирования называется мягким. Обычно декодеру сообщается квантованное напряжение с выхода демодулятора — нескольких бит квантования оказывает-

ся достаточным. Мягкое декодирование в пределе для G_A дает дополнительный энергетический выигрыш в 3 дБ.

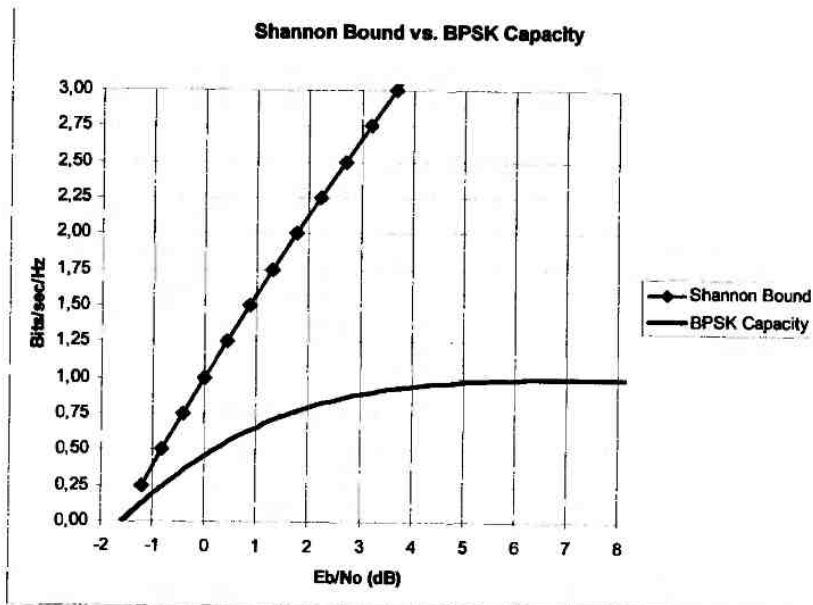


Рис. 4.4. Сравнение пропускной способности каналов: произвольного с аддитивным гауссовым шумом (Shannon Bound) и с двоичной модуляцией и корректирующим кодом в зависимости от отношения сигнал/шум на бит (BPSK Capacity)

4.3.10 Низкоплотностные коды

Низкоплотностные коды были предложены Р. Галлагером в 1960 году [27] и долгое время оставались красивой идеей, слабо реализуемой из-за вычислительной сложности. Однако в последнее время низкоплотностные коды становятся одной из достаточно модных процедур кодирования и декодирования. Ниже приводится одна из наиболее простых, но достаточно эффективных процедур реализации низкоплотностных кодов.

Метод построения проверочной матрицы

Декодирование низкоплотностных кодов осуществляется с помощью их проверочной матрицы, которая порождается следующим образом. Пусть J и $K > J$ это некоторые целые положительные числа, а P — некоторое простое число. Обозначим через E единичную матрицу размером $P \times P$. Рассмотрим исходную проверочную матрицу H_0 , которая представляет собой J рядов из K подматриц, каждая из которых получается в результате некоторой циклической перестановки столбцов единичной матрицы E . В дальнейшем совокупность строк проверочной матрицы, соответствующих некоторому ряду подматриц, будем на-

зывать полосой. Отметим, что каждый столбец любой полосы содержит ровно одну единицу. Соответственно, каждый столбец матрицы H_0 содержит J единиц, а каждая строка — K единиц.

Полученная исходная матрица H_0 задает линейный код длиной $N = P \times K$ со скоростью передачи $R \approx 1 - J/K$. Приблизительный знак равенства использован потому, что, как правило, построенная таким образом проверочная матрица имеет линейно зависимые строки. Таким образом, число информационных символов оказывается несколько большим, чем $P \times (N - J)$. При $K > P$ можно выбрать циклические перестановки столбцов единичных подматриц таким образом, что матрица H_0 будет задавать J ортогональных проверок для каждого символа кода. Будем говорить, что такая матрица обладает свойством ортогональности проверок, и будем использовать ее при порождении других матриц путем перестановки столбцов в каждой полосе исходной проверочной матрицы.

Задается некоторое целое положительное число M — количество попыток перестановок столбцов в каждой полосе, которые осуществляются последовательно для каждой полосы. При каждой очередной попытке псевдослучайно выбираются два столбца в данной полосе, которые затем меняются местами. После этого проводится проверка наличия свойства ортогональности проверок. При положительном результате этой проверки перестановка столбцов сохраняется, в противном случае — аннулируется. После проведения всех попыток перестановок во всех полосах матрицы H_0 получается некоторая псевдослучайная проверочная матрица H , обладающая свойством ортогональности проверок. Как показало проведенное моделирование низкоплотностных кодов в канале с белым аддитивным гауссовым шумом, матрица H , как правило, обеспечивает лучшие результаты, чем матрица H_0 .

Алгоритм декодирования низкоплотностных кодов

Описываемый ниже алгоритм декодирования низкоплотностных кодов основывается на использовании демодулятора с мягким решением. Таким образом, из демодулятора в декодер должны передаваться два вектора: вектор жестких двоичных решений для каждого символа $(x_1, x_2, \dots, x_N)$ и вектор надежностей символов $(E_1, E_2, \dots, E_N)$, где значение надежности i -го символа — величины E_i пропорционально логарифмическому отношению правдоподобия символа x_i . По полученному вектору жестких решений для символов кода вычисляется синдром $(s_1, s_2, \dots, s_r)$, где r — число строк проверочной матрицы.

Алгоритм декодирования состоит из заданного числа итераций. При каждой итерации выполняется последовательная обработка всех строк проверочной матрицы. Обработка строки заключается в коррекции надежностей входящих в данную строку символов кода в зависимости от значения соответствующей данной строке компоненты синдрома. При этом следует отметить, что в данном алгоритме надежности символов рассматриваются как положительные величины. Если при коррекции надежности некоторого символа его надежность стала отрицательной, жесткое значение символа меняется на противоположное, при этом его надежность меняет знак на положительный.

Опишем теперь детально процедуру обработки строки. Рассмотрим строку с номером j . Перед обработкой данной строки имеются текущие значения векторов жестких решений, надежностей и синдрома. После обработки строки выдаются модифицированные значения этих векторов. Обозначим через $n(j, k)$ позицию k -й единицы в j -й строке. Таким образом, $n(j, 1), n(j, 2), \dots, n(j, K)$ — это позиции единиц в строке с номером j . Рассмотрим набор надежностей символов, соответствующих единицам в данной строке, $E_{n(j,1)}, E_{n(j,2)}, \dots, E_{n(j,K)}$. Далее для простоты будем обозначать эти надежности как $E_1, E_2, \dots, E_K$.

Обработка каждой строки состоит из двух шагов. На первом после последовательного просмотра набора надежностей находятся позиции двух наименьших надежностей. Пусть E_{k1} — минимальная надежность, а E_{k2} — надежность, не превосходящая все остальные, кроме E_{k1} .

Выполнение второго шага зависит от значения компоненты синдрома, соответствующей данной строке. Если $s_j = 0$, то в позициях $k1$ и $k2$ устанавливаются значения надежностей $E_{k1} + E_{k2}$, в остальных позициях значения надежностей увеличиваются на E_{k1} . Если $s_j = 1$, то в позициях $k1$ и $k2$ устанавливаются значения надежностей $E_{k2} - E_{k1}$, в остальных позициях значения надежностей уменьшаются на E_{k1} . Кроме того, в этом случае изменяется жесткое значение символа, соответствующего позиции $k1$, и производится соответствующая модификация синдрома.

После обработки всех строк проверочной матрицы проводится процедура, необходимость которой не вытекает из математической сущности алгоритма, а обусловлена ограничениями на технические возможности вычислительных устройств. Речь идет о так называемой процедуре нормировки надежностей. Дело в том, что после обработки всех строк проверочной матрицы значения надежностей заметно увеличиваются. Таким образом, после нескольких итераций эти значения могут выйти за рамки допустимых для данного вычислительного устройства. Для того чтобы избежать этого нежелательного эффекта, значения всех надежностей с помощью соответствующего сдвига помещаются в некоторый заданный диапазон.

Оценка сложности алгоритма декодирования

В качестве оценки вычислительной сложности алгоритма будет рассматриваться количество элементарных арифметических операций в алгоритме декодирования. Под элементарными операциями будем понимать сложение, вычитание и сравнение целых положительных чисел, а также сложение по модулю 2 двоичных чисел и операцию сдвига целых чисел на определенное число разрядов. Сложность алгоритма определяется как суммарное количество указанных выше операций.

Для первоначального вычисления синдрома требуется не более JN сложений по модулю 2.

Для выполнения шага 1 при обработке одной строки требуется не более двух $(K - 2)$ сравнений.

Для выполнения шага 2 при обработке одной строки требуются $K+2$ операции сложения или вычитания, а также не более J сложений по модулю 2 в случае корректировки синдрома.

Для проведения процедуры нормировки надежностей достаточно N операций сравнения и $N + J$ операций сдвига.

Обозначим через $L(N)$ вычислительную сложность одной итерации для последовательного вычислительного устройства (без предварительного вычисления синдрома). Легко видеть, что

$$L(N) \leq (2(K-2) + K + 2 + J) + 2N + J.$$

Учитывая, что $\tau = \frac{NJ}{K}$ и $\frac{J}{K} \approx 1 - R$, получаем

$$L(N) \leq NJ(4 - R).$$

Пусть Q — количество итераций алгоритма, а $L(Q, N)$ — вычислительная сложность всего алгоритма при Q итерациях для последовательного вычислительного устройства. Тогда

$$L(Q, N) \leq QNJ(4 - R) + NJ = QNJ(4 - R - 1/Q).$$

Обозначим через $L(Q)$ вычислительную сложность всего алгоритма на один кодовый символ при Q итерациях для последовательного вычислительного устройства. Тогда

$$L(Q) \leq QJ(4 - R - 1/Q).$$

Данная оценка вычислительной сложности алгоритма получена для последовательного вычислительного устройства, в котором в каждый момент времени выполняется одна элементарная операция. Далее будет оцениваться вычислительная сложность алгоритма в предположении, что имеется «параллельное» вычислительное устройство, в котором в каждый момент времени может выполняться одновременно любое количество элементарных операций. Обозначим через $L^*(N)$, $L^*(Q, N)$ и $L^*(Q)$ вычислительную сложность одной итерации, сложность всего алгоритма при Q итерациях и сложность на один кодовый символ при Q итерациях соответственно для параллельного вычислительного устройства.

Легко видеть, что для параллельного вычислительного устройства сложность предварительного вычисления синдрома приблизительно равна $J \log_2 N$, а сложность процедуры нормировки $\log_2 N$.

Поскольку в любой полосе проверочной матрицы каждый столбец содержит ровно одну единицу, все строки в одной полосе могут обрабатываться одновременно. Следовательно, параллельная сложность обработки одной полосы не превышает $3K + J - 2$.

Таким образом,

$$L^*(N) \leq J(3K + J - 2) + \log_2 N,$$

$$L^*(Q, N) \leq J \log_2 N + QJ(3K + J - 2) + Q \log_2 N,$$

$$L^*(Q) \leq ((J + Q) \log_2 N + QJ(3K + J - 2))/N.$$

В качестве примера рассмотрим низкоплотностный код длиной $N = 4112$, $R = 0,75$, $J = 4$, $K = 16$. Пусть при декодировании этого кода используется 10 итераций. Тогда с помощью приведенных формул находим, что в случае

последовательного вычислительного устройства на один символ приходится 126 элементарных операций, в то время как для параллельного устройства — не более 0,6 операции на один символ.

Хотя определенное выше параллельное вычислительное устройство является практически нереализуемым, поскольку требуется одновременное выполнение неограниченного числа элементарных операций, тем не менее результат, полученный для этого устройства имеет определенный практический смысл. Дело в том, что требуемая параллельность вычислений при реализации алгоритма декодирования ограничивается количеством строк в одной полосе проверочной матрицы, равным P (в рассмотренном выше примере $P = 257$). Таким образом, для достижения результата параллельного устройства, технически реализуемое устройство должно выполнять одновременную обработку P строк в одной полосе проверочной матрицы.

Оценим теперь объем оперативной памяти, требуемый для реализации алгоритма декодирования. Для хранения векторов жестких решений, надежностей и синдрома требуется N , $2N$ и $(1 - R)N$ байт соответственно (при этом не используется упаковка компонент жестких решений и синдрома). Основным объемом оперативной памяти уходит на хранение двух матриц, содержащих позиции единиц в строках и столбцах проверочной матрицы соответственно, на что требуется $4NJ$ байт. Таким образом, общий объем составляет $N(4(J + 1) - R)$ байт. Отсюда для рассмотренного выше кода получаем около 80 кбайт.

Результаты моделирования помехоустойчивости

На рис. 4.5–4.7 иллюстрируются результаты моделирования трех низкоплотностных кодов в гауссовом канале с двоичной фазовой модуляцией. Вероятность ошибки в бите P_b обозначена BER, а вероятность обнаружения ошибки в кодовом слове обозначена BLER.

Кодовые параметры (8384,7863,521) (seed=1237 shuffles=4000)

E_b/N_0	4,8	4,90	5,00	5,10	5,20
(BER)	2,8838E-03	1,1984E-03	3,4850E-04	1,0058E-04	1,4909E-05
(BLER)	3,4247E-01	1,6949E-01	5,5866E-02	1,7379E-02	2,7813E-03

E_b/N_0	5,30	5,40	5,50	5,6
(BER)	1,7564E-06	2,0692E-07	2,4377E-08	2,8718E-09
(BLER)	3,7139E-04	4,9592E-05	6,6221E-06	8,8425E-07

N	K	R			
8384	7863	0,937858			
$L_d(Q, N)$	$L_d^*(Q, N)$	$L_c(N)$	$L_c^*(N)$	V_m , Кбайт	V_{rom} , Кбайт
993384	7819,151	128019,5	7863	159,817	512,0779
$[L_d(Q)]$	$[L_d^*(Q)]$	$[L_c]$	$[L_c^*]$		
118,4857	0,932628	15,2695	0,937858		

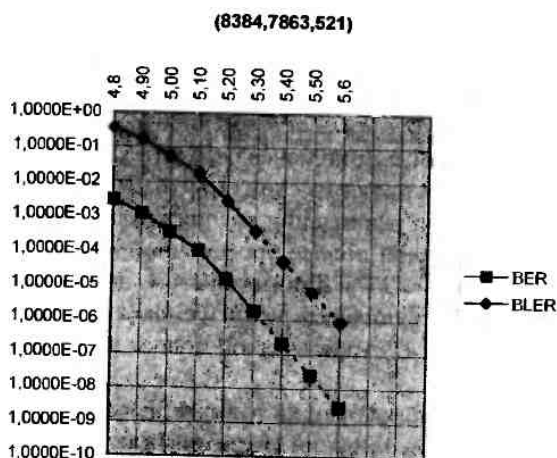


Рис. 4.5. Вероятность ошибки в бите P_b (BER) и вероятность обнаружения ошибки в кодовом слове (BLER) в зависимости от отношения сигнал/шум на бит для низкоплотного кода в гауссовом канале с двучисловой фазовой модуляцией при кодовых параметрах (8384,7863,521) (seed=1237 shuffles=4000)

4.3.11 Сверточные коды

Первый непрерывный рекуррентный код был предложен в 1955 году советским ученым Л.М. Финком [95]. Однако спустя 4 года (в 1959 году) «вновь открытый» рекуррентный код был назван по имени его западного автора кодом Хегельбергера [77]. Если a_k — информационные символы, а b_k — проверочные, то кодовая последовательность имеет вид $a_1, b_1, a_2, b_2, \dots, a_k, b_k, a_{k+1}, b_{k+1}, \dots$. Информационные символы определяются передаваемым сообщением, а проверочные формируются по правилу $b_k = a_{k-s} \oplus a_{k+s+1} \pmod{2}$, где s — произвольное целое число, называемое шагом кода ($s = 0, 1, 2$). Очевидно, что при ошибочном приеме некоторого проверочного символа b_i кодовое соотношение в принятой последовательности не будет выполнено для $i = k$. В случае же ошибочного приема информационного символа a_i кодовое соотношение не будет выполняться при двух значениях k . В принятой кодовой последовательности для каждого b_k проверяется кодовое соотношение. Если оно оказалось невыполненным при двух значениях k , разделенных величиной $2s + 1$, то информативный элемент a_{k+s+1} должен быть заменен на противоположный. Очевидно, что избыточность такого кода равна $1/2$. Так, если $s = 0$, обеспечивается правильное декодирование, когда между двумя ошибочно принятыми символами имеется не менее трех правильно принятых символов.

Кодовые параметры (8224,7199,1025) (seed=1237 shuffles=8000)

E_b/N_0	3,60	3,70	3,80	3,90	4,00	4,10
(BER)	2,4071E-02	2,0366E-02	1,5011E-02	8,8418E-03	4,6465E-03	1,8007E-03
(BLER)	9,9010E-01	9,4340E-01	8,1301E-01	5,5866E-01	3,4364E-01	1,5152E-01

E_b/N_0	4,20	4,30	4,40	4,50	4,60	4,7
(BER)	4,4340E-04	5,0334E-05	6,3057E-06	4,5647E-07	3,3044E-08	2,3920E-09
(BLER)	3,9185E-02	5,0103E-03	7,2245E-04	5,4000E-05	4,0363E-06	3,0169E-07

N	K	R			
8224	7199	0,875365			
$L_d(Q, N)$	$L_d^*(Q, N)$	$L_c(N)$	$L_c^*(N)$	V_m , Кбайт	V_{rom} , Кбайт
994984	3946,05	230593	7199	157,281	922,3719
$[L_d(Q)]$	$[L_d^*(Q)]$	$[L_c]$	$[L_c^*]$		
120,9854	0,479821	28,03903	0,875365		

(8224,7199,1025)

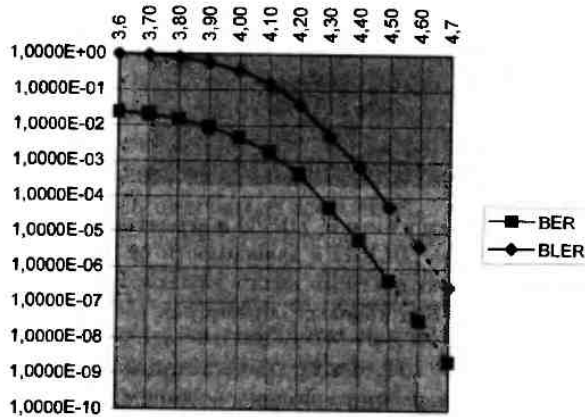


Рис. 4.6. Вероятность ошибки в бите P_b (BER) и вероятность обнаружения ошибки в кодовом слове (BLER) в зависимости от отношения сигнал/шум на бит для низкоплотного кода в гауссовом канале с двучной фазовой модуляцией при кодовых параметрах (8224,7199,1025) (seed=1237 shuffles=8000)

Кодовые параметры (8336,6255,2081) (seed=1237 shuffles=4000)

E_b/N_0	2,8	2,90	3,00	3,10	3,20	3,30
(BER)	4,5229E-02	3,8264E-02	2,8797E-02	1,5087E-02	6,1945E-03	1,6505E-03
(BLER)	9,9010E-01	9,9010E-01	9,1743E-01	5,4945E-01	2,5907E-01	7,9808E-02

E_b/N_0	3,40	3,50	3,60	3,70	3,80	3,9
(BER)	2,4494E-04	1,8341E-05	9,6529E-07	5,0803E-08	2,6738E-09	1,4072E-10
(BLER)	1,2903E-02	1,0507E-03	5,7460E-05	3,1423E-06	1,7185E-07	9,3978E-09

N	K	R			
8336	6255	0,75036			
$L_d(Q, N)$	$L_d^*(Q, N)$	$L_c(N)$	$L_c^*(N)$	V_m , Кбайт	V_{rom} , Кбайт
1050216	2017,661	406770,5	6255	160,465	1627,082
$[L_d(Q)]$	$[L_d^*(Q)]$	$[L_c]$	$[L_c^*]$		
125,9856	0,242042	48,79684	0,75036		

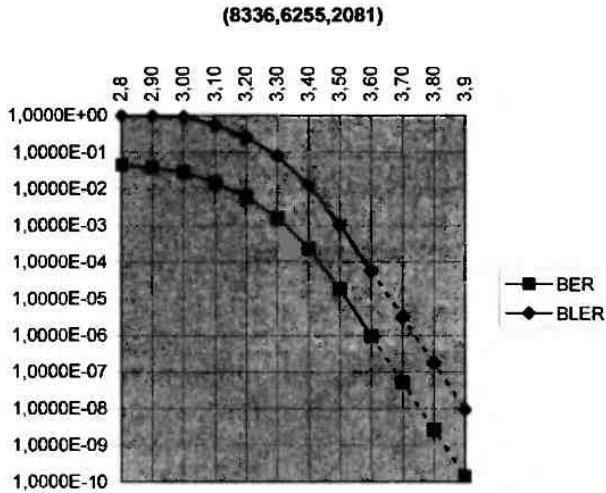


Рис. 4.7. Вероятность ошибки в бите P_b (BER) и вероятность обнаружения ошибки в кодовом слове (BLER) в зависимости от отношения сигнал/шум на бит для низкоплотного кода в гауссовом канале с двоичной фазовой модуляцией при кодовых параметрах (8336,6255,2081)(seed=1237 shuffles=4000)

Структура сверточного двоичного кодера в общем виде имеет k регистров, куда за один кодовый такт входят k символов, и n сумматоров, связанных с определенными разрядами регистров. Связи j -го сумматора по $(mod 2)$ описываются путем задания j -й порождающей последовательности $\bar{g}_i = (g_{i0}, g_{i1}, \dots, g_{i(m-1)})$, $i = 1, \dots, k$, где m — длина регистра, а наличие коэффициента в последовательности соответствует наличию соединения. Типичные параметры сверточных кодов: $k, n = 1, 2, \dots, 8$; $R = k/n = 1/4, \dots, 7/8$; $m = 2, \dots, 10$ [43]. Наиболее часто на практике применяются сверточные коды со скоростью $R = 1/2$. Сверточный код удобно задавать посредством порождающих многочленов, определяемых видом последовательностей $\bar{g}_i = (g_{i0}, g_{i1}, \dots, g_{i(m-1)})$, подобно тому как это делается для линейных блочных циклических кодов. Порождающие многочлены полностью определяют структуру кодера двоичного сверточного кода. Выходные кодовые символы можно представить в виде свертки последовательности информационных символов и порождающих многочленов кода, задающих линейные рекуррентные правила кодирования. Порождающие многочлены могут быть объединены в матрицу размером $k \times n$, называемую порождающей матрицей из многочленов. Используя представление сверточного кода, с помощью порождающих коэффициентов производящих многочленов, записанных в двоичной (или восьмеричной) форме.

Очевидно, для рекуррентных и, в частности, для сверточных кодов понятие кодового слова не имеет смысла, так как кодовые символы вычисляются по текущему блоку последних информационных символов для каждого такта работы кодера. Поэтому подобные коды называют также цепными, или скользящими.

Вообще говоря, сверточные коды можно рассматривать как обобщение блочных кодов, так как за каждый такт работы кодера по текущему блоку последних информационных символов формируется определенный блок выходных кодовых символов. Однако блоки выходных кодовых символов сверточного кода, формируемые в следующие друг за другом такты работы кодера, являются функционально зависимыми, тогда как при блоковом кодировании зависимость между соседними блоками кодовых символов (кодowymi словами) отсутствует.

Сверточный кодер как конечный автомат с памятью описывают диаграммой состояний. Диаграмма состояний представляет собой направленный граф, вершины которого отождествляются с возможными состояниями кодера, а ребра, помеченные стрелками, указывают возможные переходы между состояниями. Над каждым из ребер записывают кодовые символы, порождаемые кодером при соответствующем переходе из состояния в состояние.

Определенная последовательность кодируемых информационных символов задает конкретную последовательность смены состояний, и при этом порождаются кодовые символы, записанные над соответствующими ребрами, соединяющими состояния кодера на диаграмме состояний.

Рассматриваемую диаграмму состояний можно развернуть во времени, при этом получается так называемая решетчатая диаграмма. На ней принято, что штриховые линии (ветви) соответствуют переходам, происходящим при приходе информационного символа 1, а сплошные линии (ветви) — информационного символа 0. Из решетчатой диаграммы видно, что ее структура после окончания «переходного процесса» в кодере становится повторяющейся. Важное значение решетчатого представления состоит в том, что с ростом числа входных символов число вершин в решетке не растет, а остается равным 2^{m-1} , где m — число ячеек в регистре сдвига. Решетчатая диаграмма показывает все разрешенные пути, по которым может продвигаться кодер при кодировании.

Сверточные коды являются непрерывными и характеризуются многими минимальными расстояниями, определяемыми длинами начальных сегментов кодовых последовательностей. Число символов в принятой для обработки длине сегмента L определяет на приемной стороне число ячеек в декодирующем устройстве.

Это число символов, которое декодер должен хранить в памяти для обработки принимаемой кодовой последовательности, называется шириной окна декодирования. Если ставится цель обнаружения и исправления как можно большего числа конфигураций ошибок, то в общем случае увеличение ширины окна декодирования всегда приводит к улучшению характеристик, однако в конце концов происходит насыщение.

В соответствии с различной длиной L обрабатываемых в декодере сегментов минимальное расстояние Хемминга для любых пар кодовых слов называется L — минимальным свободным расстоянием сверточного кода и обозначается d_L . Если L достаточно велико, то это просто минимальное свободное расстояние $d_{\min}$. Очевидно, потенциально корректирующая способность сверточного кода тем выше, чем больше его минимальное свободное расстояние.

Практическая реализация сверточных кодов со скоростями $R = k/n$ встречает затруднения, особенно в случае больших скоростей передачи. Упрощение

алгоритма обработки может быть получено при выборе кода с $R = 1/n$ и «выкалывании», или удалении некоторых символов в выходной последовательности для получения кода с $R = k/n$. Такие коды называются перфорированными.

В настоящее время используется три основных метода декодирования сверточных кодов: пороговое [67], аналогичное мажоритарному методу декодирования блоковых кодов, последовательное [26, 37] и декодирование по алгоритму Витерби [12]. Наиболее простыми в реализации являются алгоритмы мажоритарного декодирования как блоковых, так и сверточных кодов. Сложность реализации декодеров растет практически пропорционально полной длине кодового ограничения. Декодеры достаточно просты при исправлении ошибок невысокой кратности. Однако дальнейшее увеличение кратности исправляемых ошибок приводит к значительному усложнению схемного построения декодеров, которое не оправдывается возрастанием величины ЭВК. Наибольшую сложность имеют декодеры Витерби, объем вычислений (сложность) которых возрастает экспоненциально с ростом длины кодового ограничения. При использовании алгоритма Витерби увеличение длины кодового ограничения на единицу более чем вдвое увеличивает объем декодера, но дает прирост ЭВК, равный 0,4–0,5 дБ [4]. Метод последовательного декодирования является методом вероятностного декодирования, при котором число операций, необходимых для декодирования одного символа, является случайной величиной.

При пороговом декодировании вычисляются синдромы, затем эти синдромы или последовательности, полученные посредством линейного преобразования синдромов, подаются на входы порогового элемента, где путем «голосования» (мажоритарный метод) и сравнения его результатов с порогом выписится решение о значении декодируемого символа. Основное достоинство этого метода декодирования — простота реализации. Однако он не полностью реализует потенциальные корректирующие способности сверточного кода. Кроме того, не все сверточные коды могут быть декодированы этим методом. Чтобы сверточный код допускал декодирование пороговым методом, он должен обладать свойством ортогональности.

При последовательном декодировании число операций, которое должен выполнить декодер, для того чтобы декодировать один символ, изменяется в зависимости от уровня шумов в канале. Число операций при последовательном декодировании является функцией скорости передачи и шумов в канале. При всех скоростях передачи, меньших определенной скорости, число операций при декодировании оказывается небольшим. Последовательный декодер строится по схеме, позволяющей проводить вычисления со средней скоростью в несколько раз большей скорости передачи символов, и включает в свой состав буферное запоминающее устройство, предназначенное для хранения поступающих данных при повышении уровня шумов в канале. В случае, если число возникших в канале ошибок превысит корректирующую способность кода или переполнится буфер, возникают ошибки декодирования.

Идея алгоритма Витерби состоит в том, что в декодере воспроизводят все возможные пути последовательных изменений состояния сигнала, сопоставляя получаемые при этом кодовые символы с принятыми аналогами по каналу связи, и на основе анализа ошибок между принятыми и требуемыми символами

определяют оптимальный путь. Декодирование по методу Витерби, по существу, представляет собой алгоритм поиска наилучшего, максимально правдоподобного пути на графе — решеточной диаграмме кода. При декодировании с мягким решением в качестве оптимального пути выбирают путь с максимальной суммарной метрикой, что соответствует максимальной накопленной доверительной вероятности.

Важным достоинством декодера Витерби является то, что, когда в результате воздействия шумов в канале связи или по другим причинам при декодировании сделана ошибка в выборе пути на решетчатой диаграмме кода, т. е. выбран неправильный путь, за несколько тактов, в течение которых могут происходить ошибки при декодировании, декодер вновь выходит на правильный путь.

Эффективность сверточных кодов зависит от выбранного кода и характеристики канала передачи. Итоговую оценку производят по величине коэффициента ошибок при заданном отношении сигнал/шум в канале и типе канала. Поскольку сверточные коды хорошо исправляют одиночные ошибки, но чувствительны к пакетам ошибок, то их обычно применяют вместе с перемежением передаваемых по каналу символов или в качестве внутренних в каскадном коде.

Анализ эффективности сверточных кодов показывает [4], что применение коротких сверточных кодов, декодируемых по алгоритму Витерби с мягким решением, позволяет получить ЭВК порядка 4–6 дБ. Переход к жесткому решению снижает ЭВК примерно на 2 дБ. Квантование выхода демодулятора на четыре уровня снижает ЭВК на 0,7–0,8 дБ, а квантование на восемь уровней — на 0,25 дБ. Обычно ограничиваются квантованием на восемь уровней, используя практически полностью возможности мягкого решения. Также следует отметить, что ЭВК в случае сверточных кодов растет достаточно медленно с уменьшением вероятности ошибки декодирования P_b , что также оправдывает их применение в каскадной паре в качестве внутренних кодов.

4.3.12 Турбокоды

Хотя турбокоды и были введены французским ученым Клодом Берри (Claude Berrou) сравнительно недавно — в 1993 году [127] и, следовательно, представляют собой новый тип кодов для исправления ошибок, они содержат «давно забытые» идеи начального этапа теории кодирования. Первая из этих идей — коды-произведения или итеративные коды, которыми фактически и являются турбокоды. Вторая идея — многократное декодирование, когда результат предыдущего декодирования используется на последующих шагах. Вторая идея использовалась, например, Р. Галлагером при декодировании низкоплотностных кодов. Однако эти старые идеи в турбокодах обрели «второе дыхание». Уже в первой работе по турбокодам [127] была практически продемонстрирована возможность получения значения вероятности ошибки в бите $P_b = 10^{-5}$ для ФМ2 в канале с аддитивным белым гауссовым шумом при отношении сигнал/шум на бит всего 0,7 дБ, что лишь на 0,5 дБ больше теоретического предела.

Турбокоды есть и сверточные, и блочные. Исторически раньше появились сверточные турбокоды, а позже блочные. Сверточные турбокоды удобнее для малых относительных скоростей кода, а блочные для больших.

Сверточные турбокоды

Принцип построения кодера турбокода достаточно прост [64] и показан на рис. 4.8. Из структуры кодера видно, что турбокод представляет собой систематический непрерывный код, в котором проверочные символы генерируются двумя кодерами составных рекурсивных сверточных кодов (РСК), причем информационная последовательность подается в кодер первого РСК (РСК1) непосредственно, а в кодер второго РСК (РСК2) — через устройство псевдослучайного перемежения. Схема «выкалывания» проверочных символов применяется для регулирования общей скорости турбокода.

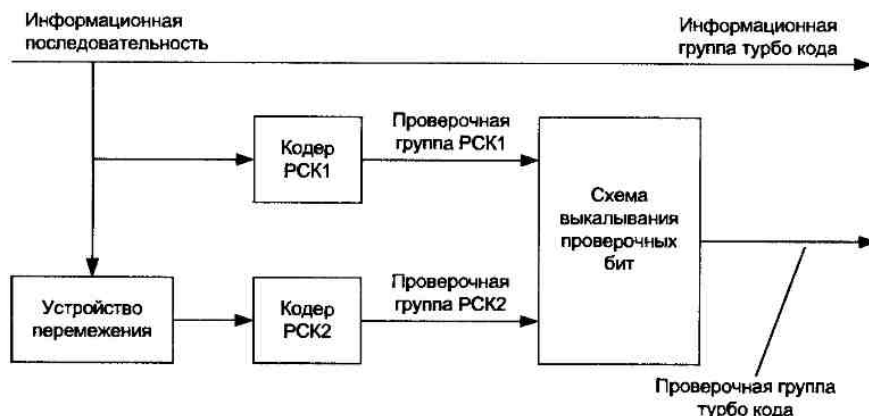


Рис. 4.8. Структура кодера турбокода

Если при подаче определенной информационной последовательности на вход кодера РСК1 вес его проверочной последовательности оказывается малым, то перемеженная версия этой информационной последовательности, подаваемая на вход кодера РСК2, с высокой вероятностью приведет к генерации проверочной последовательности большого веса. Если какая-либо комбинация ошибок не может быть исправлена одним РСК, то это почти наверняка будет сделано с помощью другого РСК, и наоборот. Заметьте, что при использовании в составе турбокода нерекурсивной формы сверточных кодов с такой же корректирующей способностью, выигрыш от кодирования оказывается намного меньше. Это происходит потому, что вес выходной последовательности сверточных кодов в нерекурсивной форме слабо зависит от вида входной информационной последовательности.

Число составных кодов в турбокоде может быть и больше двух.

Для декодирования турбокодов в настоящее время повсеместно применяется концепция так называемого итеративного декодирования, сущность которой можно раскрыть, рассматривая структуру итеративного декодера турбокода (рис. 4.9).

Итеративный декодер образован последовательным соединением декодеров двух элементарных кодов (РСК1 и РСК2), так называемых декодеров с мягкими входным и выходным сигналом (Soft-In Soft-Out — SISO).

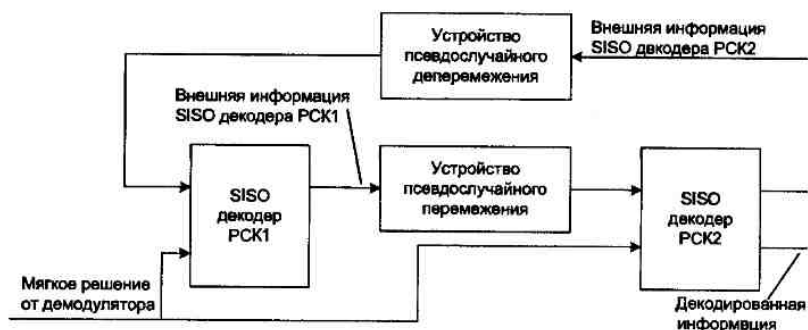


Рис. 4.9. Структура итеративного декодера турбокода

В процессе декодирования турбокода элементарные SISO декодеры обмениваются друг с другом внешней информацией, с каждой итерацией улучшая окончательное решение в смысле снижения вероятности ошибки в бите в декодированной информационной последовательности (одна итерация включает в себя декодирование PCK1 и PCK2). Однако уже после первой итерации внешняя информация, подаваемая на вход декодера PCK1 по цепи обратной связи, оказывается коррелированной с информацией, получаемой из мягких решений демодулятора для проверочных символов PCK1. Поэтому улучшение окончательного решения с каждой итерацией становится меньше и, таким образом, величина вероятности ошибки на бит, достигаемая декодированием по этому методу, стремится к определенному пределу. Окончательное (жесткое) решение о передаваемых информационных битах принимается после завершения последней итерации декодером PCK2 и подается на его отдельный выход. Методы построения SISO элементарных декодеров практически сводятся к использованию алгоритмов декодирования элементарных кодов, способных вырабатывать мягкие выходные решения о передаваемых информационных символах.

В качестве примера рассмотрим характеристики помехоустойчивости конструкции на основе ФМ2 и турбокода со следующими параметрами: $R = 1/2$, составляющие коды со скоростями $2/3$ с 16 состояниями с полиномами в восьмеричном виде 33 и 31. На рис. 4.10 приведены зависимости вероятности ошибки в бите (P_b) от отношения сигнал/шум на бит E_b/N_0 для случая использования указанного выше турбокода совместно с ФМ2 в канале связи с АБГШ и независимыми ошибками (без квантования выходного сигнала демодулятора) для нескольких значений длин информационного пакета K и при двух значениях количества итераций для каждой длины пакета (цифры около каждой кривой) [64]. Из рисунка следует, что ЭВК турбокода существенно увеличивается с ростом длины информационного пакета. Так, при увеличении длины пакета в 32 раза (4096 вместо 128) ЭВК увеличивается более чем на 2 дБ при одинаковом числе итераций (для $P_b = 10^{-5}$). При больших длинах пакета увеличение числа итераций также оказывается энергетически выгодным. Для $K = 128$ увеличение числа итераций в два раза приводит к возрастанию ЭВК в среднем только на 0,1–0,2 дБ, в то время как при $K = 4096$ это увеличение уже составляет около 0,3 дБ.

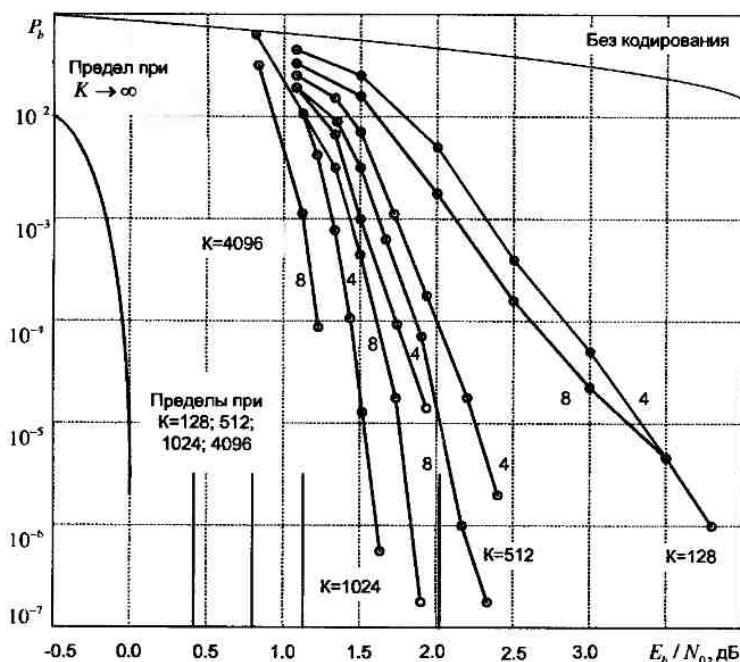


Рис. 4.10. Зависимость вероятности ошибки в бите от отношения сигнал/шум на бит (см. текст)

Блочные турбокоды

Блочные турбокоды являются кодами-произведениями или итеративными кодами. В качестве составляющих кодов используются либо простейшие коды с проверкой на четность, либо расширенные коды Хемминга с общей проверкой на четность. Коды могут быть не только двумерными, но и трехмерными. В некоторых случаях возможны комбинации с низкоплотными кодами. В качестве алгоритмов декодирования составляющих кодов используются специализированные алгоритмы мягкого декодирования, выдающие метрику каждого бита на выход декодера (SISO). Декодирование, как и в случае сверточных турбокодов, осуществляется несколько раз.

На рис. 4.11 и 4.12 показаны результаты сравнительной помехоустойчивости блочных турбокодов и каскадной пары — внутренний сверточный код и внешний код РС. Результаты взяты с сайта компании АНА (www.aha.com) [61], производителя микросхем для этих кодовых конструкций. На рис. 4.11 показаны результаты для релейского канала с замираниями, а на рис. 4.12 — для канала с аддитивным гауссовым шумом.

Из рисунков видно, что двумерные турбокоды с составляющими расширенными кодами Хемминга (128,120,4) приводят к результатам, даже в гауссовском

канале дающим дополнительные 4 дБ ЭВК по сравнению с каскадной парой «сверточный код с декодированием по Витерби и код РС». Также важно, что результирующие относительные скорости блочного турбокода существенно выше, чем у сверточных.

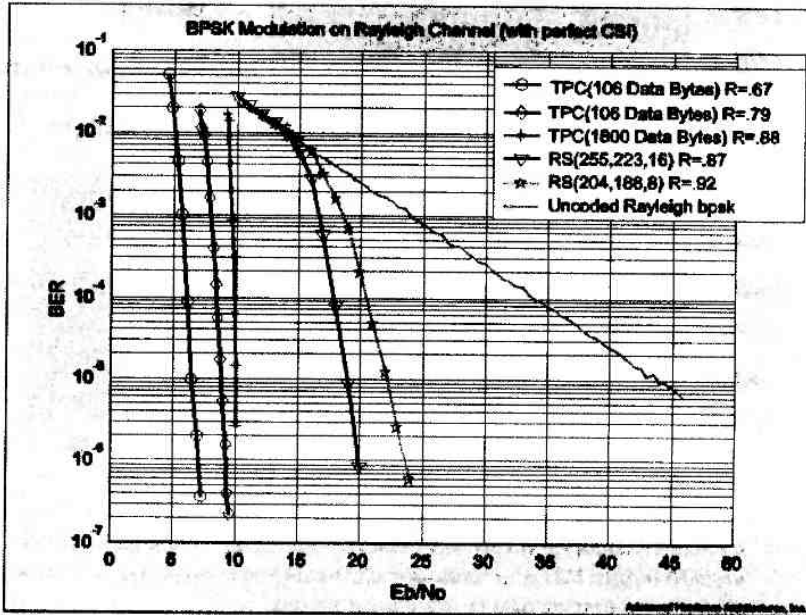


Рис. 4.11. Сравнительная помехоустойчивость блочных турбокодов (TPC) и каскадной пары (RS) для релейского канала с замираниями в зависимости от отношения сигнал/шум на бит

4.3.13 Обобщенные каскадные коды

Обобщенные каскадные коды были предложены советскими учеными Э.Л. Блохом и В.В. Зябловым для линейного случая [7] и В.А. Зиновьевым для нелинейного [38]. Обобщенный каскадный код в трактовке Э.Л. Блоха и В.В. Зяблова представляет собой следующее.

Двоичное слово $\bar{a}$ длины $n_1 n_2$ представляется в виде последовательности двоичных векторов $\bar{a}_j$, $j = 1, \dots, n_2$, длины n_1 , т. е. $\bar{a} = (\bar{a}_1, \bar{a}_2, \dots, \bar{a}_{n_2})$.

Исходными данными для описания обобщенного каскадного кода являются:

- n_2 квадратных двоичных матриц $H_0^{(j)}$, $j = 1, \dots, n_2$, порядка n_1 ;
- $m+1$ групповых над $GF(2^{a_i})$, $i = 1, \dots, m+1$ кодов (n_2, b_i) . При этом $a_1 + a_2 + \dots + a_m + a_{m+1} = n_1$.

Коды (n_2, b_i) над полем $GF(2^{a_i})$ называются кодами второй ступени, или внешними кодами.

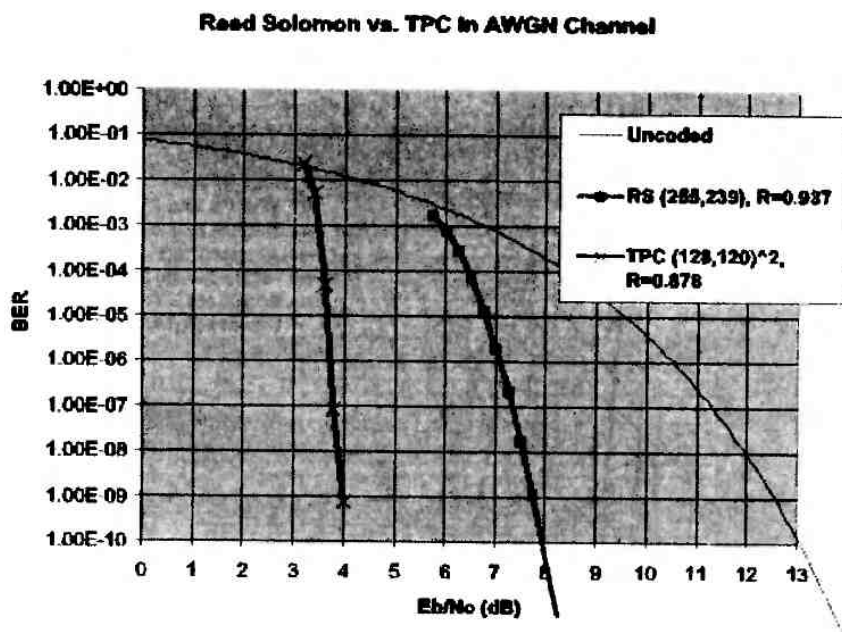


Рис. 4.12. Сравнительная помехоустойчивость блочных турбокодов (TPC) и каскадной пары (RS) для канала с аддитивным гауссовым шумом в зависимости от отношения сигнал/шум на бит

Можно ввести линейное отображение векторов $\bar{a}_j$

$$\bar{a}_j H_0^j = (\bar{\gamma}_{1j}, \bar{\gamma}_{2j}, \dots, \bar{\gamma}_{m+1,j}), j = 1, \dots, n_2,$$

где $\bar{\gamma}_{ij}$ — двоичный вектор длины a_i . Трактруя векторы $\bar{\gamma}_{ij}$ как элементы поля $GF(2^{a_i})$, для каждого вектора $\bar{a}$ определим векторы $\bar{\gamma}_i = (\bar{\gamma}_{i1}, \bar{\gamma}_{i2}, \dots, \bar{\gamma}_{im_2})$, $i = 1, \dots, m+1$.

Тогда обобщенный каскадный код определяется так.

Двоичное слово $\bar{a}$ длины $n = n_1 n_2$ является кодовым словом обобщенного каскадного кода порядка m тогда и только тогда, когда все связанные со словом $\bar{a}$ векторы $\bar{\gamma}_i$, $i = 1, \dots, m+1$ представляют собой кодовые слова соответствующих i -ых кодов второй ступени. Такой обобщенный каскадный код является линейным.

Таким образом, образующими кодами обобщенного каскадного кода являются $m+1$ кодов первой ступени, или внутренних кодов $(n_1, (a_1 + a_2 + a_i), d_{1i})$, и столько же кодов второй ступени, или внешних кодов (n_2, b_i, d_{2i}) , над полем $GF(2^{a_i})$, где $i = 1, \dots, m+1$. Фактически, слово линейного обобщенного каскадного кода является поразрядной суммой по модулю 2 $m+1$ слов обычных каскадных кодов Форни с внутренним кодом $(n_1, (a_1 + a_2 + a_i), d_{1i})$ и внешним кодом (n_2, b_i, d_{2i}) над полем $GF(2^{a_i})$ (у данного кода внешний код кодирует только соответствующие a_i символов внутреннего кода), где $i = 1, \dots, m+1$. Обычные каскадные коды Форни называются обобщенными каскадными кодами нулевого порядка.

На основе данной трактовки легко убедиться, что результирующий линейный обобщенный каскадный код имеет следующие параметры:

$$\left\{ \begin{array}{l} n = n_1 n_2, \\ k = \sum_{i=1}^{m+1} a_i b_i, \\ d_{\min} = \min_{i=1, \dots, m+1} (d_{1i} d_{2i}) \end{array} \right\}.$$

Существенным требованием при построении и существовании линейного обобщенного каскадного кода с данными параметрами является необходимость вложенности внутренних кодов одного в другой, т. е.

$$(n_1, a_1, d_{1,1}) \subset (n_1, (a_1 + a_2), d_{1,2}) \subset \dots \subset (n_1, (a_1 + a_2 + \dots + a_m), d_{1,m}) \subset (n_1, (a_1 + \dots + a_m + a_{m+1}), d_{1,m+1}).$$

Естественно, что для минимальных расстояний внутренних кодов выполняется условие $d_{1,1} \geq d_{1,2} \geq \dots \geq d_{1,m} \geq d_{1,m+1}$. Для максимизации минимального расстояния обобщенного каскадного кода естественно выбирать минимальные расстояния внешних кодов d_{2i} таким образом, чтобы выполнялось условие $d_{\min} = d_{11}d_{21} = d_{12}d_{22} = \dots = d_{1,m+1}d_{2,m+1}$.

Самый простой алгоритм декодирования состоит из $m+1$ шагов. На каждом шаге осуществляется попеременное декодирование i -ым внутренним, затем i -ым внешним кодом и коррекция исправленных символов в принятом слове. Затем происходит переход к $i-1$ шагу, $i = m+1, \dots, 1$. Суммарная вероятность ошибки декодирования складывается из $m+1$ соответствующих вероятностей ошибок на каждом шаге. Расстояния внешних кодов d_{2i} выбираются из условия равенства составляющих вероятностей ошибки декодирования. Практика показывает, что это происходит при условии большей защищенности начальных «слоев» декодирования $d_{1,m+1}d_{2,m+1} \geq d_{1,m}d_{2,m} \geq \dots \geq d_{12}d_{22} \geq d_{11}d_{21}$. Переход к обобщенным каскадным кодам от каскадного кода нулевого порядка приводит к дополнительному ЭВК более 1 дБ [39]. В качестве внешних в линейных обобщенных кодах обычно используются РС коды.

Ключевым при построении обобщенных каскадных кодов m -го порядка является требование наличия системы из $(m+1)$ -го вложенного внутреннего кода. При этом, если не требовать линейности от системы внутренних кодов, получается естественное обобщение на нелинейный случай. Такие нелинейные обобщенные коды будут использованы в следующей главе при построении сигнально-кодовых конструкций.

ГЛАВА 5

Системы модуляции и сигнально-кодовые конструкции

5.1 Модуляция как перенос сигнала по спектру

Под модуляцией в технике связи подразумевают изменение параметров несущего сигнала в соответствии с параметрами информационного сигнала. Такое наложение информационного сигнала на несущий может быть необходимо, например, потому, что последний хуже распространяется в конкретной среде, являющейся каналом связи. Однако перенос спектра это одна и не самая важная функция модуляции. Гораздо важнее то, что модуляция обеспечивает перенос информации с «цифрового» кодового описания на «непрерывное» сигнальное, т. е. переводит дискретную метрику Хемминга в непрерывную метрику Евклида.

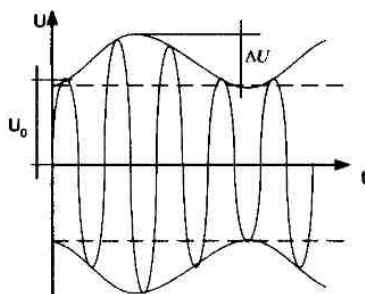


Рис. 5.1. Амплитудно-модулированный гармонический сигнал

Наибольшее распространение в качестве несущих получили гармонический сигнал и периодическая последовательность импульсов. На рис. 5.1 показан результат модуляции по амплитуде несущего сигнала $U_0 = \cos(\omega_0 t + \varphi_0)$ также гармоническим сигналом $\Delta U \cos(\Omega t + \psi)$. Отношение $m = \Delta U/U_0$ носит название коэффициента модуляции. Очевидно, что $0 \leq m \leq 1$.

Аналитическое выражение для данного сигнала имеет вид:

$$U(t) = [U_0 + \Delta U \cos(\Omega t + \psi)] \cos(\omega_0 t + \varphi_0).$$

В результате его преобразования получается:

$$U(t) = U_0 \cos(\omega_0 t + \varphi_0) + \frac{\Delta U}{2} \cos[(\omega_0 + \Omega) + (\varphi_0 + \psi)] + \\ + \frac{\Delta U}{2} \cos[(\omega_0 - \Omega) + (\varphi_0 - \psi)].$$

На рис. 5.2 показан спектр результирующего сигнала при модуляции гармоническим сигналом, периодическим сигналом сложной формы и произвольным сигналом. Из рис. 5.2 видно, что спектр модулированного сигнала имеет три составляющие — с несущей частотой ω_0 и так называемыми «боковыми» частотами, отличающимися от несущей на величину $+\Omega$ и $-\Omega$. Поскольку этот результат распространяется на любую частотную составляющую произвольного модулирующего сигнала, можно легко определить качественный вид спектра гармоник, модулируемой периодическим сигналом сложной формы, аperiodическим сигналом произвольной формы со спектром в диапазоне от Ω_n до Ω_b .

Из рис. 5.2 следует важный результат — с помощью модуляции спектр модулирующего сигнала можно перенести в любую необходимую область частотного диапазона.

До сих пор речь шла лишь об амплитудной модуляции (АМ) гармонического сигнала. Наряду с амплитудой в качестве изменяемого параметра несущей можно использовать также частоту и фазу (ЧМ и ФМ). Не вдаваясь в подробности анализа спектра ЧМ и ФМ сигналов, отметим лишь, что такие сигналы также включают несущую и полосы боковых частот. При этом последние шире, чем при АМ.

Еще один распространенный вариант — модуляция последовательности импульсов. При амплитудно-импульсной модуляции (АИМ), частотно-импульсной (ЧИМ), фазово-импульсной (ФИМ) и широтно-импульсной модуляции (ШИМ) пропорционально амплитуде модулирующего сигнала меняются соответствующие названию параметры несущего сигнала.

В дальнейшем будем считать, что результирующий модулированный сигнал $U(t) = A(t) \cos(\omega_0 t + \varphi(t))$ является узкополосным, т.е. ширина спектра сигнала существенно меньше его несущей частоты $\Delta\omega \ll \omega_0$. Тогда вся информация, имеющаяся в сигнале, заложена в его комплексной огибающей $A(t)e^{j\varphi(t)}$, а это значит, что на так называемой сигнальной плоскости можно показать все изменения сигнала. В каждый момент времени сигнал отображается на сигнальной плоскости точкой с амплитудой и фазой.

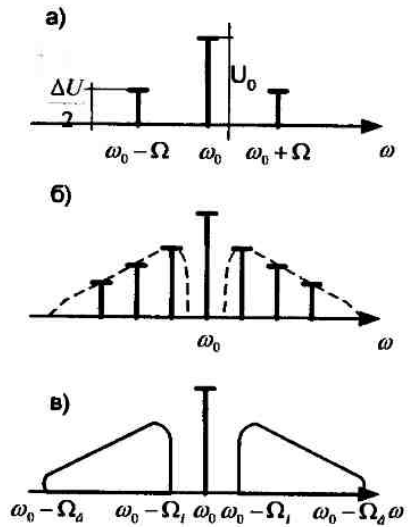


Рис. 5.2. Спектр результирующего сигнала при модуляции гармоническим сигналом (а), периодическим сигналом сложной формы (б) и произвольным сигналом (в)

5.2 Дискретная модуляция

В случае дискретной модуляции сигнал $U(t) = A(t) \cos(\omega_0 t + \varphi(t))$ в каждый момент времени принимает одно из набора $Q = 2^q$ дискретных значений. Все множество $Q = 2^q$ сигналов называется сигнальным алфавитом. Как и коррек-

тирующий код, сигнальный алфавит характеризуется двумя основными параметрами — по скорости и по помехоустойчивости.

По скорости ансамбль сигналов характеризуется скоростью в битах на измерение канала: $R = q/2$, где q — число бит, которыми могут быть перенумерованы сигналы. По помехоустойчивости ансамбль сигналов характеризуется квадратом нормированного по средней мощности ансамбля евклидова расстояния Δ^2 .

Самый простой вид дискретной модуляции — фазовая (ФМ). Все сигналы ФМ лежат на комплексной плоскости на одной окружности через равные промежутки. Легко убедиться, что в этом случае квадрат нормированного по средней мощности ансамбля евклидова расстояния $\Delta^2 = 4 \sin^2(\pi/2^q)$. Например, в случае ФМ8 $\Delta^2 = 0,576$. Достоинством ФМ является постоянная амплитуда сигнала, что приводит к ее использованию в каналах, где имеется нелинейное преобразование амплитуды передаваемого сигнала.

Однако фазовая модуляция обладает одним существенным недостатком — не используется сигнальная плоскость внутри круга, что приводит к уменьшению Δ^2 . Исходя из этого в большинстве реализованных систем используется наиболее общий вид модуляции — амплитудно-фазовая (АФМ). Самая простая и легко реализуемая разновидность АФМ — так называемая квадратурная амплитудная модуляция (КАМ), когда сигнал располагается в узлах равномерной квадратной решетки. В случае четного q построение очевидно. В случае нечетного q ансамбль сигналов КАМ строится «выкалыванием» точек через одну построчно из ансамбля сигналов КАМ с четным q и с q на единицу большим. В общем случае квадрат нормированного по средней мощности ансамбля евклидова расстояния есть

$$\Delta^2 = \begin{cases} 6/(2^q - 1), & q - \text{четное,} \\ 6/(2^q - 1/2), & q - \text{нечетное} \end{cases}.$$

Так, например, в случае КАМ16 имеем $\Delta^2 = 0,4$.

Для большого числа сигналов ансамбля в случае фазовой модуляции ФМ приближенно имеем $\Delta^2 \approx \pi^2/2^{2(q-1)}$, а в случае КАМ $\Delta^2 \approx 3/2^{(q-1)}$. Это значит, что в случае увеличения ансамбля сигналов вдвое помехоустойчивость фазовой модуляции ФМ падает в четыре раза, т. е. на 6 дБ, а квадратурной амплитудной модуляции КАМ — в два раза, т. е. на 3 дБ.

Отдельной проблемой является нумерация сигналов ФМ или КАМ. Одним из наиболее распространенных вариантов является нумерация сигналов кодом Грея, когда ближайшие по евклидову расстоянию сигналы отличаются номерами с расстоянием Хемминга, равным 1. В случае КАМ выбирается код Грея и по горизонтали, и по вертикали. Другим видом нумерации является обычная двоичная нумерация.

5.3 Сигнально-кодовые конструкции (СКК) в гауссовом канале

До сих пор мы оценивали эффективность систем с двоичной модуляцией и кодированием при помощи энергетического выигрыша кодирования (ЭВК). Такая

оценка не учитывает частотной эффективности системы связи. Пусть R [бит/с], есть скорость передачи информации конкретной системы связи, $\Delta\omega$ [Гц], есть ширина спектра сигнала, а E_b/N_0 [дБ], есть отношение энергии одного бита к спектральной плотности шума. В соответствии с [4] определим частотную эффективность системы как $\gamma = R/\Delta\omega$ [бит/с/Гц].

Эффективность системы определим как отношение скорости передачи к пропускной способности канала $\eta = R/C$, где C есть пропускная способность, определяемая по формуле Шеннона (см. рис. 4.4):

$$C = \Delta\omega \log_2(1 + P_c/P_{\text{ш}}),$$

где $P_c/P_{\text{ш}}$ — отношение сигнал/шум, вычисляемое по формуле

$$P_c/P_{\text{ш}} = (E_b/N_0)(R/\Delta\omega) = \gamma(E_b/N_0).$$

Тогда эффективность системы может быть определена в следующем виде:

$$\eta = R/(\Delta\omega \log_2(1 + P_c/P_{\text{ш}})) = \gamma/\log_2(1 + P_c/P_{\text{ш}}).$$

Максимальное значение коэффициент эффективности системы принимает при $\eta_{\text{max}} = 1$. Тогда для оптимальной системы (на границе пропускной способности) получим:

$$\gamma = \log_2(1 + \gamma E_b/N_0) \text{ или } 2^\gamma = 1 + \gamma E_b/N_0.$$

Оптимальные обменные соотношения между помехоустойчивостью E_b/N_0 (энергетическая эффективность) и скоростью передачи γ (частотная эффективность), известные как граница Шеннона, окончательно представляются в следующем виде:

$$E_b/N_0 = (2^\gamma - 1)/\gamma.$$

Таким образом, можно оценивать эффективность любой реальной системы передачи как точку в осях E_b/N_0 и γ , ограниченных кривой Шеннона сверху. Следует помнить, что кривая Шеннона показывается для вероятности ошибки, стремящейся к нулю, а точка конкретной системы фиксируется для конкретной вероятности ошибки в бите. Расстояния от конкретной точки системы до кривой Шеннона по осям абсцисс и ординат есть проигрыш пропускной способности по скорости и помехоустойчивости. Также видно, что для увеличения частотной эффективности γ на 1 бит/с/Гц необходимо увеличить энергетическую эффективность на 3 дБ. Из этого следует, что КАМ с ростом числа сигналов ведет себя по отношению к пропускной способности Шеннона одинаково (не приближается и не удаляется), ФМ с ростом числа сигналов отдаляется от пропускной способности — асимптотический угол наклона в два раза больше.

Анализ эффективности систем показывает, что использование многопозиционной модуляции хотя и позволяет повысить скорость передачи по сравнению с двоичной модуляцией, но не позволяет приблизиться к пропускной способности ни по частотной, ни по энергетической эффективности. Использование же корректирующих кодов вместе с двоичной модуляцией позволяет приблизиться к пропускной способности для частотной эффективности, меньшей единицы.

Как же приблизиться к пропускной способности в области с высокой частотной эффективностью? Здесь прослеживается полная аналогия с корректирующими кодами и дискретными каналами. Совершенно понятно, что использова-

ния сигналов на двумерной сигнальной плоскости недостаточно для достижения пропускной способности. Необходимо передавать информацию многомерными сигналами, где размерность сигнала $N \rightarrow \infty$. При этом ансамбль M сигналов строится таким образом, что все сигналы являются точками в N -мерном пространстве Евклида. Сигналы необходимо выбирать таким образом, чтобы вокруг каждого сигнала можно было бы очертить непересекающиеся N -мерные сферы максимального радиуса.

Тогда, по аналогии с корректирующими кодами ансамбль таких сигналов будем обозначать $(N, K, D_{\min}^2)$, где N — размерность сигнала, $K = \log_2 M$ — число бит, переносимых сигналом, $D_{\min}^2$ — нормированный по средней мощности ансамбля сигналов $P_{\text{ср}}$ квадрат минимального расстояния Евклида. Скорость такого ансамбля есть $R = K/N$ [бит/изм.].

В этих обозначениях сигналы ФМ или КАМ из 2^q точек принимают вид $(2, q, \Delta^2)$. В этих обозначениях, хотя и прослеживаются аналогии с корректирующими кодами, есть некоторые отличия. В отличие от ненормированного расстояния по Хеммингу $d_{\min}$ квадрат минимального расстояния Евклида $D_{\min}^2$ нормируется по средней мощности ансамбля сигналов $P_{\text{ср}}$. Из этого следует, что при построении многомерных ансамблей сигналов надо стремиться не к росту $D_{\min}^2$, а минимизировать его падение.

Рассмотрим пример, иллюстрирующий выигрыш при увеличении размерности сигнала.

Пусть имеется сигнал ФМ4 с параметрами (2,2,2). Скорость есть $R = 1$ бит/изм. Все 4 сигнала ФМ4 на сигнальной плоскости отображаются точками на круге через равные промежутки

$$\bar{x}_2 = (x_1, x_2) = \{(+1, 0), (0, +1), (-1, 0), (0, -1)\}.$$

Разложим сигнал ФМ4 на два подмножества ФМ2 —

$$\{(+1, 0), (-1, 0)\} \quad \text{и} \quad \{(0, +1), (0, -1)\}.$$

Будем формировать четырехмерный сигнал $\bar{x}_4 = (x_1, x_2, x_3, x_4)$ как последовательный набор двух сигналов ФМ4, причем первый сигнал выбирается произвольно, а второй берется из того же самого подмножества, что и первый. Всего возможно 8 четырехмерных сигналов следующего вида:

$$\bar{x}_4 = \left\{ \begin{array}{l} (+1, 0, +1, 0), (+1, 0, -1, 0), (0, +1, 0, +1), (0, +1, 0, -1), \\ (-1, 0, +1, 0), (-1, 0, -1, 0), (0, -1, 0, +1), (0, -1, 0, -1) \end{array} \right\}.$$

Видно, что реальный квадрат минимального расстояния Евклида для этого ансамбля по сравнению с ФМ4 вырос в два раза, а нормированный по мощности 4-мерного сигнала остался без изменения. В наших обозначениях параметры этого ансамбля есть (4,3,2), $R = 0,75$ бит/изм. Если сравнивать этот ансамбль с ФМ2, то при том же расстоянии Евклида скорость в полтора раза больше. Точки этого ансамбля выбраны в 4-мерном пространстве, что и обеспечивает этот выигрыш.

Таким образом, мы убедились, что задача построения оптимального ансамбля сигналов сводится к задаче оптимальной упаковки шаров в N -мерном пространстве Евклида и нормировании потом по средней мощности сигнала. Последняя

задача, хотя и является традиционной математической задачей, мало подходит для данного случая из-за отсутствия непереборных алгоритмов декодирования таких конструкций. Исходя из наличия хорошо развитой теории корректирующих кодов, было предложено решать вопросы оптимальных упаковок в пространстве Евклида размерности N при помощи сигнально-кодовых конструкций (СКК). СКК фактически является каскадным или обобщенным каскадным кодом, где в качестве внутреннего кода используется простой ансамбль двумерных сигналов, а в качестве внешнего — корректирующий код [29, 38, 39, 78]. Возможно использование как блоковых, так и сверточных корректирующих кодов.

5.4 Описание блоковых СКК в гауссовом канале

Как сказано выше, блоковая СКК является нелинейным каскадным кодом с внутренней системой сигналов и внешним корректирующим кодом. Ввиду различия метрик системы сигналов и корректирующих кодов возможно несколько способов согласования метрик и отсюда — несколько СКК [39].

Конструкция 1 состоит из одного внутреннего ансамбля сигналов и одного внешнего корректирующего кода. Внутренний ансамбль 2^q двумерных сигналов (например, ФМ или КАМ) имеет параметры $(2, q, \Delta^2)$. Внешним является корректирующий код $(N/2, k, d_{\min})$ с элементами из поля $GF(2^q)$. Тогда СКК имеет следующие параметры:

N — размерность,

$K = qk$ — число информационных символов,

$D_{\min}^2 = 2\Delta^2 d_{\min}/N$ — квадрат минимального расстояния Евклида, нормированный по средней мощности.

В данной конструкции нумерация сигналов в ансамбле модуляции не важна, так как q -й корректирующий код не использует тот факт, что разные сигналы находятся друг от друга на разных расстояниях. Прием СКК состоит в последовательном выполнении демодуляции ансамбля сигналов и декодировании корректирующим кодом.

Конструкция 2 также состоит из одного внутреннего ансамбля сигналов и одного внешнего корректирующего кода. Внутренний ансамбль сигналов такой же, как в конструкции 1, а внешний корректирующий код является двоичным с параметрами $(qN/2, k, d_{\min})$. В [39] показано, что параметры конструкции следующие:

N — размерность,

$K = k$ — число информационных символов,

$D_{\min}^2 = 2\Delta^2 d_{\min}/N$ — квадрат минимального расстояния Евклида, нормированный по средней мощности.

Для обеспечения указанных параметров существенна нумерация сигналов во внутреннем ансамбле сигналов — она должна быть кодом Грея. Для кода Грея и в случае ФМ, и в случае КАМ выполняется условие, что два двумерных сигнала, отличающиеся в m битах, имеют разницу в расстоянии Евклида не менее $m\Delta^2$. Тогда при доказательстве минимального расстояния СКК можно использовать

тот факт, что у двух слов на минимальном расстоянии друг от друга различные биты разбросаны по разным сигналам.

Прием данной СКК также состоит в последовательном выполнении демодуляции ансамбля сигналов и декодирования корректирующим кодом. Отличие состоит в том, что в случае мягкого декодирования внешним кодом вырабатываются специальные метрики для каждого бита в отдельности в одном и том же сигнале.

Конструкция 3 является обобщенным каскадным кодом $(q - 1)$ -го порядка и состоит из q пар внутренних и внешних кодов. В качестве внутренних кодов используется система вложенных ансамблей сигналов.

В случае ансамблей ФМ имеем вложенную цепочку ансамблей $\text{ФМ2} \subset \text{ФМ4} \subset \text{ФМ8} \subset \dots \subset \text{ФМ2}^q$. Параметры модуляции на каждом шаге соответствуют описанным выше.

В случае ансамблей КАМ вложенная цепочка ансамблей принимает вид

$$(\text{КАМ2})' \subset (\text{КАМ4})' \subset (\text{КАМ8})' \subset \dots \subset \text{КАМ2}^q.$$

Здесь только исходная модуляция является обычной КАМ модуляцией. Все остальные ансамбли получаются один из другого «прореживанием» точек в два раза. Таким образом на каждом шаге прореживания квадрат минимального расстояния Евклида увеличивается ровно в два раза, правда, немного смещается центр ансамбля.

Внешние корректирующие коды имеют параметры $(N/2, k_i, d_i)$, $i = \overline{1, q}$.

Тогда СКК имеет следующие параметры:

N — размерность,

$K = \sum_{i=1}^q k_i$ — число информационных символов,

$D_{\min}^2 = \min_{i=\overline{1, q}} (2\Delta_i^2 d_i / N)$ — квадрат минимального расстояния Евклида, нор-

мированный по средней мощности, а Δ_i^2 — квадрат нормированного расстояния ансамбля сигналов на i -м шаге вложенности.

Прием такой СКК аналогичен декодированию обобщенных каскадных кодов и состоит из q шагов. На i -м шаге происходит демодуляция сигналов в алфавите из 2^{q+1-i} сигналов, декодирование i -м внешним кодом и коррекция ансамблей сигналов по результатам декодирования (выбор соответствующего ансамбля из 2^{q-i} сигналов для следующего шага).

Возможна комбинация этих конструкций, когда в конструкции 3 на i -м шаге происходит разбиение на q_i ансамблей сигналов и используется конструкция 1 или 2.

Приведем пример построения конструкций конкретной размерности.

В качестве ансамбля сигналов используем ансамбль квадратурной амплитудной модуляции КАМ16. Размерность всех сравниваемых конструкций выберем $N = 32$, а квадрат минимального расстояния Евклида, нормированный по средней мощности, $D_{\min}^2 = 0, 2$.

Тогда в конструкции 1 внешним является код РС над полем $GF(16)$ с параметрами $(16, 9, 8)$. СКК содержит $4 \times 9 = 36$ информационных символов и имеет скорость $R = 1, 125$ бит/изм.

В конструкции 2 выберем в качестве внешнего код Рида-Маллера 3-го порядка (64,42,8). СКК содержит 42 информационных символа и имеет скорость $R = 1,3125$ бит/изм.

В конструкции 3 выберем в качестве внешних коды (16,5,8), (16,11,4), (16,15,2) и (16,16,1). СКК содержит $5 + 11 + 15 + 16 = 47$ информационных символов и имеет скорость $R = 1,46875$ бит/изм.

Как видно, для данного примера эффективность конструкций возрастает с номером конструкции. Часто конструкция 2 оказывается лучше конструкции 3. Конструкция 1 обычно им проигрывает. В [38] приводятся подробные примеры построения СКК с различными длинами и числом информационных символов.

5.5 Описание сверточных СКК в гауссовом канале

Для сверточных кодов в принципе возможно построение таких же трех конструкций, как и в случае блочных кодов. Однако все эти конструкции прежде всего эффективны для кодов регулярной структуры. Наиболее же эффективные и часто употребляемые сверточные коды являются кодами с достаточно коротким кодовым ограничением, строятся методами перебора и декодируются также перебором по алгоритму Витерби. Наверное, благодаря сказанному, наибольшее распространение для сверточных СКК получил отдельный подход, начало которому положила работа Г. Унгербека [208].

Суть подхода в следующем. Сверточный код, как и ранее, выбирается методами перебора из всего класса кодов данной структуры. На ребрах, в отличие от двоичного случая, выбираются либо непосредственно сигналы, либо номера подмножеств в разложении сигналов на достаточно большое число подмножеств. Также ребрам могут соответствовать не двумерные, а четырехмерные сигналы. Таким образом, строится оптимальная сверточная конструкция заданной сложности (с заданным числом узлов решетчатой диаграммы). Достигаются ЭВК порядка 5–6 дБ при том же R , что и в безызбыточной конструкции.

5.6 Модель канала с межсимвольной интерференцией (МСИ)

В реальных частотно-ограниченных каналах связи помимо аддитивного шума возникает межсимвольная интерференция (МСИ), вызванная памятью каналов. Отклик канала на последовательность входных сигналов вызывает взаимное наложение сигналов на выходе канала. Если нормировать по мощности амплитудно-частотную характеристику канала, то можно сказать, что МСИ приводит к значительному изменению расстояний между сигналами на выходе канала и, что особенно важно, к уменьшению минимального расстояния между ними.

При синтезе сигналов и кодов для каналов с МСИ этот эффект, как правило, не учитывается, т. е. в качестве входных сигналов выбираются такие, которые согласованы с идеальным каналом без МСИ. Однако МСИ стремятся учитывать

при синтезе оптимального приемника (декодера). Широко известным решением такого рода является алгоритм Витерби и его модификация, учитывающая сверточное кодирование [12, 28].

Рассмотрим подход к кодированию в каналах с МСИ, основанный на синтезе таких сигнально-кодовых конструкций, которые учитывают «деформацию» пространства сигналов при передаче по реальному каналу [39, 40]. Основой этого подхода является показанная ниже возможность преобразования каналов с МСИ в совокупность гауссовых каналов без памяти, т. е. без МСИ, но отличающихся один от другого скалярным коэффициентом передачи или отношением сигнал/шум. В 90-х годах прошлого века данный подход получил название ортогональной частотной модуляции или ортогонального частотного мультиплексирования — Orthogonal Frequency Division Modulation (OFDM).

Широкий класс реальных каналов может быть представлен в виде линейной модели, состоящей из линейного фильтра и сумматора с аддитивным шумом [12, 28]. Вход и выход такого канала связаны выражением

$$\tilde{Z}(t) = \int_0^{\infty} h(\tau)Z(t - \tau)\alpha\tau + \gamma(t),$$

где $h(\tau)$ — импульсная реакция канала (отклик на δ -функцию), $\gamma(\tau)$ — аддитивный шум. Импульсная реакция канала, аддитивный шум и входной сигнал $Z(t)$ в общем случае являются комплексными. Импульсная реакция канала $h(t)$ связана с передаточной функцией канала $K(\omega)$ преобразованием Фурье

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} K(\omega)e^{j\omega t} \alpha\omega.$$

Положим, что канал является низкочастотным, т. е. $K(\omega)$ существенно отличается от нуля в интервале, серединой которого является $\omega = 0$. Для любых линейных методов модуляции всегда может быть получен низкочастотный эквивалент канала. Канал также предполагается физически возможным, т. е. каузальным: $|h(t)| = 0, \forall t < 0$. Функция $h(t)$ принадлежит пространству функций $L_2(0, \infty)$, из чего следует следующее утверждение.

Существует точка на оси времени $t_{rp}(\varepsilon)$, такая, что

$$\int_{t_{rp}(\varepsilon)}^{\infty} |h(t)|^2 \alpha t < \varepsilon, \text{ где } \varepsilon > 0.$$

Это позволяет аппроксимировать канал системой с конечной импульсной реакцией — в дальнейшем будем считать, что $h(t) = 0, \forall t > t_{rp}$. Цифровой сигнал на входе канала может быть представлен в виде

$$Z(t) = \sum_q Z_q \delta(t - qT), \quad q = \overline{-\infty, \infty},$$

где Z_q — комплексный отсчет сигнала на входе канала,

$\delta(t)$ — дельта функция, T — тактовый интервал.

Сигнал на выходе канала может быть представлен как

$$\tilde{Z}(t) = \sum_q Z_q h(t - qT) + \gamma(t).$$

Этот сигнал в приемнике дискретизируется по времени с частотой $f_T = 1/T$ в моменты времени $kT + Q$, где Q — начальная фаза стробирования. В результате этой операции получается:

$$\tilde{Z}(kT + Q) = \sum_q Z_q h(kT + Q - qT) + \gamma(kT + Q).$$

Обозначим через

$$\begin{aligned} \tilde{Z}_k &= \tilde{Z}(kT + Q), \quad k - q = n, \\ h_n &= h(kT + Q - qT), \quad \gamma_k = \gamma(kT + Q). \end{aligned}$$

Тогда получаем выражение, связывающее последовательность отсчетов на входе канала Z_k , $k = -\infty, \infty$ с последовательностью на его выходе:

$$\tilde{Z}_k = \sum_n h_n Z_{k-n} + \gamma_k.$$

В дальнейшем будем считать, что аддитивный шум является белым и гауссовым (АБГШ). Это значит, что γ_k — комплексная гауссова величина: $\gamma_k = \varsigma_k + j\iota_k$, причем $E(\gamma_k) = 0$,

$$E(|\gamma_k|^2) = \sigma_{\text{ш}}^2, \quad \forall k, \quad E(\gamma_k^* \gamma_l) = 0, \quad \forall k \neq l, \quad E(\varsigma_k \iota_k) = 0, \quad \forall k.$$

Величина $\sigma_{\text{ш}}^2$ представляет собой мощность АБГШ на входе приемника.

Описанная модель представляет собой общепринятую модель канала с межсимвольной интерференцией [12].

Последовательность отсчетов импульсной реакции $\{h_n\}$ будет конечной, поскольку сама импульсная реакция конечна. Пусть $h_0 \neq 0$, $h_{L-1} \neq 0$ соответственно первый и последний отсчеты импульсной реакции канала $h(t)$. При этом параметр L определяет память канала, поскольку каждый отсчет сигнала на входе представляет собой линейную комбинацию L переданных отсчетов.

Импульсной характеристикой гауссова канала дискретного времени называется последовательность отсчетов $\{h_n\}_{n=0}^{L-1}$.

Перепишем сигнал на выходе канала с учетом конечности импульсной характеристики

$$\tilde{Z}_k = \sum_{n=0}^{L-1} h_n \sqrt{b^2 - 4ac} Z_{k-n} + \gamma_k.$$

Именно это выражение и будет использовано в дальнейшем для описания канала, коэффициенты h_n могут быть как постоянными, так и медленно меняться во времени, что соответствует радиоканалам с переменными параметрами.

Временной характеристикой канала дискретного времени является импульсная характеристика $\{h_n\}_{n=0}^{L-1}$, а частотной — периодическая функция вида

$$K_g(\omega) = \sum_{n=0}^{L-1} h_n e^{-j\omega nT} = \frac{1}{T} \sum_q K(\omega - q \frac{2\pi}{T}).$$

Передающей функцией в полосе Найквиста называется фрагмент (один период) $K_g(\omega)$ на интервале $[-\frac{\pi}{T}, \frac{\pi}{T}]$.

Название «полоса Найквиста» связано с критерием Найквиста, определяющим условия отсутствия МСИ.

Для отсутствия ($h_n = 0, \forall n \neq 0, h_0 = 1$) необходимо и достаточно выполнение равенства

$$\frac{1}{T} \sum_q K(\omega - q \frac{2\pi}{T}) = 1, \forall \omega.$$

Особенностью рассмотренных ниже методов передачи дискретных сообщений в канале с МСИ является то, что отсчеты сигнала поступают на вход канала не непрерывным потоком, а блоками конечной длины (N отсчетов), отделенными один от другого защитными интервалами, обеспечивающими отсутствие интерференции между блоками. В течение защитного интервала сигнал на входе канала тождественно равен нулю.

Если длина (число отсчетов) защитного интервала L_0 больше или равна памяти канала $L - 1$, то блоки на выходе канала перекрываются. Таким образом, условие $L_0 \geq L$ является условием отсутствия межблочной интерференции. Длина блока на выходе больше, чем на входе, и равна $N_0 = N + L - 1$.

Переход от непрерывной последовательности отсчетов на входе канала к последовательности изолированных блоков позволяет воспользоваться векторным описанием канала с МСИ. Пусть $\bar{Z}_N^T = (Z_0, Z_1, \dots, Z_{N-1})$ — вектор отсчетов сигнала на входе канала, $\tilde{\bar{Z}}_{N_0}^T = (\tilde{Z}_0, \tilde{Z}_1, \dots, \tilde{Z}_{N_0-1})$ — вектор отсчетов сигнала на выходе канала, $\bar{\Delta}_{N_0}^T = (\gamma_0, \gamma_1, \dots, \gamma_{N_0-1})$ — вектор отсчетов АБГШ.

Тогда вход и выход канала связаны выражением

$$\tilde{\bar{Z}}_{N_0} = K_{N_0 N} \bar{Z}_N + \bar{\Delta}_{N_0}, \quad (5.1)$$

где $K_{N_0 N}$ — матрица канала ($N_0 \times N$), элементами которой являются компоненты весовой последовательности $\{h_n\}_{n=0}^{L-1}$

$$K_{N_0 N} = \begin{bmatrix} h_0 & & & & & \\ h_1 & h_0 & . & . & . & . \\ . & h_1 & . & . & . & h_0 \\ . & . & . & . & . & h_1 \\ h_{L-1} & . & . & . & . & . \\ . & h_{L-1} & . & . & . & . \\ . & . & . & . & . & h_{L-1} \end{bmatrix}.$$

Столбцы ленточной матрицы $K_{N_0 N}$ содержат все компоненты импульсной характеристики и отличаются один от другого сдвигом ненулевых элементов, причем ранг этой матрицы равен N .

Гауссовым каналом с МСИ (ГКМСИ) будем называть такой канал дискретного времени с конечной памятью, вход и выход которого связаны выражением (5.1), а блоки сигнала на выходе канала не перекрываются (не интерферируют): $L_0 \geq L - 1$.

Импульсную характеристику ГКМСИ и его передаточную функцию в полюсе Найквиста будем считать нормированными:

$$\sum_{n=0}^{L-1} |h_n| = \frac{T}{2\pi} \int_{-\pi/T}^{\pi/T} |K_g(\omega)|^2 d\omega = 1.$$

Введение защитных интервалов между блоками приводит к тому, что часть времени канал для передачи сообщений не используется. Это уменьшает скорость передачи в N_0/N раз. Относительная скорость определяется выражением

$$\rho_0 = \frac{N}{N_0} = \frac{1}{1 + L_0/N}.$$

Однако с увеличением размера блока N потери в скорости, связанные с введением защитных интервалов, могут быть сделаны сколь угодно малыми.

Все рассмотренные ниже конструкции сигналов для ГКМСИ получены в предположении, что ограничена средняя мощность сигнала на входе канала.

Средней мощностью сигнала на входе ГКМСИ называется величина

$$P_{\text{вх}}(\cdot)^* = \frac{1}{N_0} S_p(\bar{Z}_N \bar{Z}_N^*),$$

где $(\cdot)^*$ означает транспортирование и комплексное сопряжение, $S_p(\cdot)$ — оператор следа.

Допустимая средняя мощность на входе канала $P_1 \geq P_{\text{вх}}$ определяет с учетом нормировки импульсной характеристики и передаточной функции канала отношение сигнал/шум $S/N_{\text{ш}} \triangleq P_1/\sigma_{\text{ш}}^2$.

5.7 Преобразование канала с МСИ в параллельные каналы без памяти

Решение задачи построения эффективных СКК для ГКМСИ основано на замене последнего эквивалентным ему каналом, представляющим собой множество параллельных гауссовых каналов без памяти (ГКБП). Ниже рассмотрены два способа преобразования ГКМСИ в совокупность параллельных ГКБП. Первый из этих способов обеспечивает строгую эквивалентность исходного и преобразованного каналов, однако на практике использован быть не может; второй — связан с некоторыми энергетическими потерями, но существенно менее сложен в реализации [39, 40].

Гауссовым каналом без памяти будем называть канал, вход и выход которого связаны между собой выражением $y = ax + e$, где x, y соответственно вход и выход ГКБП, $e = \zeta + j\iota$ — не зависящий от x комплексный гауссовый шум с независимыми составляющими ζ и ι , имеющими одинаковое распределение $N(0, \sigma_{\text{ш}}/\sqrt{2})$, a — некоторая константа.

Преобразование ГКМСИ с помощью сингулярного разложения матрицы канала

Можно доказать [39, 40] следующее утверждение.

ГКМСИ с N -мерными блоками на входе и N_0 -мерными блоками на выходе эквивалентен N параллельным ГКБП, выходы и входы которых связаны выра-

жением

$$\tilde{Z}_i = \sqrt{\rho_i} \tilde{Z}_i + e_i, \quad i = \overline{0, N-1},$$

где e_i — независимые по i комплексные гауссовы случайные величины, ρ_i — сингулярные значения матрицы канала K_{N_0N} (собственные значения матрицы $K_{N_0N}^* K_{N_0N}$).

Доказательство осуществляется с помощью представления матрицы ГКМСИ K_{N_0N} в виде разложения по сингулярным значениям

$$K_{N_0N} = \sum_{j=0}^{N-1} \sqrt{\rho_i} \bar{q}_i \bar{t}_i^* = Q_{N_0N} R_N^{1/2} T_N^*,$$

где Q_{N_0N} — матрица $(N_0 \times N)$: $Q_{N_0N} = (\bar{q}_0, \bar{q}_1, \dots, \bar{q}_{N-1})$, T_N — унитарная матрица $(N_0 \times N)$: $T_N = (t_0, t_1, \dots, t_{N-1})$, R_N — диагональная матрица: $R_N = \text{diag} \{\rho_0, \rho_1, \dots, \rho_N\}$; числа ρ_i , $i = \overline{0, N-1}$ являются сингулярными значениями матрицы K_{N_0N} , они совпадают с собственными значениями матрицы $K_{N_0N}^* K_{N_0N}$ и ненулевыми собственными значениями матрицы $K_{N_0N} K_{N_0N}^*$. Поскольку

$$\text{rank } K_{N_0N} = \text{rank } K_{N_0N}^* K_{N_0N} = \text{rank } K_{N_0N} K_{N_0N}^* = N, \quad \rho_i > 0, \quad \forall i = \overline{0, N-1}.$$

Последовательность преобразований ГКМСИ в совокупность параллельных независимых ГКБП состоит в следующем. В передатчике вектор сообщения $\tilde{Z}_N$ подвергается унитарному преобразованию T_N , и полученный сигнал $\tilde{Z}_N$ поступает в канал. В приемнике принятый сигнал $\tilde{Z}_{N_0}$ подвергается преобразованию $Q_{N_0N}^*$, матрица которого имеет ортонормированные столбцы $\bar{q}_i$, в результате чего получается сигнал на выходе эквивалентного канала $\tilde{\tilde{Z}}_N$, связанный с переданными выражениями:

$$\tilde{\tilde{Z}}_N = R_N^{1/2} \tilde{Z}_N + \bar{E}_N, \quad \text{где} \quad \tilde{\tilde{Z}}_N = Q_{N_0N}^* \tilde{Z}_{N_0}.$$

МСИ, действующая в исходном канале, проявляется в преобразованном канале в различии коэффициентов передачи $\sqrt{\rho_i}$ каждого из ГКБП. Заметим, что поскольку матрица $K_{N_0N}^* K_{N_0N}$ — эрмитова, то все ее собственные значения ρ_i — вещественные.

Полученные в результате преобразования ГКБП отличаются один от другого коэффициентом передачи $\sqrt{\rho_i}$ или, что то же самое, отношением сигнал/шум.

Строгий метод преобразования канала с МСИ с использованием матриц Q_{N_0N} и T_N является единственным и дает наилучший из возможных результатов, однако весьма сложен с практической точки зрения по следующим причинам. Во-первых, для его реализации требуется точное знание на передающей и приемной стороне импульсной характеристики канала и учет последней при вычислении матриц Q_{N_0N} и T_N и коэффициентов передачи $\sqrt{\rho_i}$. Учитывая сложность вычисления собственных значений и собственных векторов, это представляется весьма проблематичным. Во-вторых, для данных преобразований неизвестны «быстрые» алгоритмы уменьшения сложности.

Преобразование ГКМСИ с помощью замены апериодической свертки циклической

Выражение, связывающее вход и выход ГКМСИ, соответствует апериодической свертке последовательности отсчетов входного сигнала с импульсной характе-

Заметим, что преобразование свертки в циклическую непременно сопряжено с потерей части энергии полезного сигнала в защитных интервалах.

Средняя мощность сигнала P_{L_0} , теряемая при выделении вектора $\tilde{Z}_N$ на приеме, определяется величиной $P_{L_0} \leq P \frac{L_0}{N}$, где P — средняя мощность сигнала.

Выбором длины блока N доля мощности, теряемой в защитных интервалах, может быть сделана сколь угодно малой.

Такой канал будем называть гауссовым каналом с МСИ и циклической сверткой (ГКМСИ-Ц). Если C_N — циркулянт, то его собственные значения K_i являются компонентами N -точечного ДПФ импульсной характеристики ГКМСИ

$$K_i = \sum_{n=0}^{L-1} h_n e^{-j \frac{2\pi}{N} i n}, \quad i = \overline{0, N-1}.$$

Тогда получается, что ГКМСИ-Ц эквивалентен N параллельным ГКБП, входы и выходы которых связаны выражением

$$\tilde{Z}_i = K_i Z_i + e_i, \quad i = \overline{0, N-1},$$

где e_i — независимые по i комплексные гауссовы случайные величины, K_i — собственные значения циркулянта C_N .

Необходимо еще раз подчеркнуть, что полученные ГКБП неэквивалентны исходному ГКМСИ, а эквивалентны полученному из последнего ГКМСИ-Ц. Отличие ГКМСИ-Ц от ГКМСИ состоит, во-первых, в упомянутой выше потере части энергии полезного сигнала из-за стирания в приемнике сигнала, попадающего в эти интервалы, и, во-вторых, в том, что $\text{rank } C_N \leq N$, тогда как $\text{rank } K = N$.

С практической точки зрения второй способ преобразования ГКМСИ в параллельные ГКБП существенно предпочтительнее первого. Причины этого в следующем.

- Для реализации преобразования достаточно знать лишь верхнюю границу памяти канала $L_0 \geq L - 1$, поскольку унитарные преобразования на передаче U_N^* и на приеме U_N являются инвариантными, т. е. от канала независимыми.
- Сложность преобразования ГКМСИ в совокупность параллельных ГКБП определяется сложностью вычисления ДПФ, для чего могут быть использованы «быстрые» алгоритмы со сложностью $O(\log N)$ операций на один отсчет сигнала, например алгоритм Кули-Тьюки.

Таким образом, применение обратного ДПФ и циклического продолжения вектора сигнала на передаче и выделение неискаженного межблочной интерференцией N -мерного блока на приеме в сочетании с прямым ДПФ на приеме позволяет преобразовать ГКМСИ в совокупность параллельных независимых ГКБП, используя минимальные априорные сведения о канале (L_0) и с минимальной сложностью цифровой обработки.

В заключение целесообразно подчеркнуть связь собственных значений K_i циркулянта C_N с параметрами исходного непрерывного канала и полученного из него канала дискретного времени:

$$K_i = K_g(\omega_i) = \frac{1}{T} \sum_q K(\omega_i - q \frac{2\pi}{T}),$$

где $\omega_i = 2\pi i/NT$, $i = \overline{0, N-1}$, $K(\omega)$ — передаточная функция непрерывного канала, $K_g(\omega)$ — передаточная функция в полосе Найквиста.

5.8 Пропускная способность канала с МСИ

Пропускной способностью ГКМСИ называется максимум взаимной информации между входом и выходом канала по всем распределениям $\tilde{Z}_N$, удовлетворяющим ограничению на среднюю мощность сигнала на входе канала

$$\frac{1}{N_0} S_p E(\tilde{Z}_N \tilde{Z}_N^*) \leq P_1.$$

Если мощность АБГШ на выходе ГКМСИ равна σ_w^2 , а средняя мощность сигнала на его входе ограничена величиной P_1 , то пропускная способность ГКМСИ с матрицей K_{km} , сингулярные значения которой $\rho_0 \geq \rho_1 \geq \dots \geq \rho_{N-1} > 0$, равна

$$C = C = \rho_0 \frac{1}{N} \sum_{i=0}^M \frac{1}{2} \log_2 \left[\rho_i \frac{N}{M} \left(\frac{P_1}{\sigma_w^2} \times \frac{1}{\rho_0} + \frac{1}{N} \sum_{p=0}^M \frac{1}{\rho_p} \right) \right],$$

где $M \leq N-1$ — наибольшее число, для которого

$$P_M = P_1 \frac{N}{M} \times \frac{1}{\rho_0} + \sigma_w^2 \left(\frac{N}{M} \times \frac{1}{N} \sum_{p=0}^M \frac{1}{\rho_p} - \frac{1}{\rho_M} \right) > 0.$$

В предыдущем разделе было показано, что ГКМСИ может быть преобразован в канал с циклической матрицей ГКМСИ-Ц, отличающийся от исходного тем, что защитные интервалы на передаче заполняются циклическим продолжением блоков, энергия которых на приеме не используется.

Для такого канала справедливо следующее утверждение. Если мощность АБГШ на выходе ГКМСИ-Ц равна σ_w^2 , а средняя мощность сигнала на его выходе ограничена величиной P_1 , то пропускная способность ГКМСИ-Ц, собственные значения которого $|K_0| \geq |K_1| \geq \dots \geq |K_{N-1}|$, равна

$$C = \rho_0 \frac{1}{N} \sum_{i=0}^M \frac{1}{2} \log_2 \left[|K_i|^2 \frac{N}{M} \left(\frac{P_1}{\sigma_w^2} + \frac{1}{N} \sum_{p=0}^M \frac{1}{|K_p|^2} \right) \right],$$

где $M \leq N-1$ — наибольшее число, для которого

$$P_M = P_1 \frac{N}{M} + \sigma_w^2 \left(\frac{N}{M} \times \frac{1}{N} \sum_{p=0}^M \frac{1}{|K_p|^2} - \frac{1}{|K_M|^2} \right) > 0.$$

Отличие этих двух формул для P_M состоит в том, что во втором случае отсутствует множитель $1/\rho_0$, учитывающий тот факт, что мощность блока сигнала на входе ГКМСИ может быть увеличена в $1/\rho_0$, так как в защитном интервале мощность на входе равна нулю. В ГКМСИ-Ц мощность сигнала в защитном интервале равна средней мощности сигнала. Здесь еще раз уместно подчеркнуть, что полученные результаты представляют собой частные случаи хорошо известного [28] общего гауссова канала.

При $N \rightarrow \infty$ ГКМСИ и ГКМСИ-Ц обращаются в гауссовский канал дискретного времени, входом которого является стационарная случайная последо-

вательность $\{Z_K\}_{K \rightarrow -\infty}^{\infty}$, а выходом — стационарная случайная последовательность $\{\tilde{Z}_K\}_{K \rightarrow -\infty}^{\infty}$.

5.9 Построение СКК для канала с МСИ и переменными параметрами (OFDM)

В результате описанного выше преобразования ГКМСИ-Ц в совокупность независимых параллельных ГКБП было получено, что вход и выход каждого канала связаны выражением

$$\tilde{Z}_i = K_i Z_i + e_i, \quad i = \overline{0, N-1}.$$

Если исходный канал дискретного времени имеет существенную неравномерность амплитудно-частотной характеристики в полосе Найквиста, то полученные каналы могут быть весьма различны. Различие ГКБП должно учитываться при построении сигналов и СКК. Возможно очень много разных вариантов преобразования ГКМСИ-Ц в совокупность независимых параллельных ГКБП [40]. Рассмотрим два наиболее интересных из них с точки зрения реализации.

Исследуем возможность преобразования ГКМСИ-Ц в совокупность независимых параллельных одинаковых ГКБП [40].

Пусть, как и ранее, собственные значения матрицы канала упорядочены $|K_0| \geq |K_1| \geq \dots \geq |K_{N-1}|$. Пусть M_1 таково, что $|K_{M_1-1}| > 0$, $|K_{M_1}| = |K_{M_1+1}| = \dots = |K_{N-1}| = 0$, $0 < M_1 \leq N-1$. Иначе говоря: $\text{rank } C_N = M_1$, где C_N — матрица ГКМСИ-Ц. Положим, что для всех $0 < i \leq M_1-1$ $Z_i = \frac{1}{|K_i|} \tilde{N}_i$, $\tilde{N}_i = b_i \tilde{Z}_i$, $b_i = e^{-j \arg K_i}$.

Тогда получаем, что $\tilde{N}_i = \tilde{N}_i + \tilde{e}_i$, где $\tilde{e}_i = e^{-j \arg K_i} e_i$.

Поскольку умножение комплексной гауссовой случайной величины с независимыми действительной и мнимой частями e_i на «поворачивающий» множитель $e^{-j \arg K_i}$ не изменяет статистику шума, то вместо обозначения $\tilde{e}_i$ будем использовать e_i . Имеем:

$$\tilde{N}_i = \tilde{N}_i + e_i, \quad i = \overline{0, M_1-1}.$$

Тогда справедливо следующее утверждение.

Предысказание на передаче $Z_i = \frac{1}{|K_i|} \tilde{N}_i$ и коррекция на приеме $\tilde{N}_i = b_i \tilde{Z}_i$ преобразуют ГКБП, эквивалентные ГКМСИ-Ц, в M_1 независимых одинаковых ГКБП, вход и выход которых связаны выражением $M_1 = \text{rank } C_N$, C_N — матрица ГКМСИ-Ц.

Следующий вариант учитывает различие в каналах, но группирует их по различным «подмножествам». Рассмотрим случай, когда в параллельных ГКБП с предысказаниями используются различные алфавиты КАМ, но с одним и тем же минимальным расстоянием Евклида Δ , не зависящим от номера ГКБП i . Необходимость рассмотрения этого варианта объясняется возможностью построения на его основе эффективных сигналов и сигнально-кодовых конструкций.

Пусть $0 < \mu_1 < \mu_2 < \mu_3 < \dots < \mu_Q \leq M_1$ — некоторое разбиение последовательности номеров ГКБП. Положим, что в ГКБП с номерами от μ_{j-1} до $\mu_j - 1$ ($\mu_0 = 0$) используется алфавит КАМ с 2^{q_j} символами, $1 \leq j \leq Q$, причем

$q_1 > q_2 > \dots > q_Q \geq 1$. Это означает, что алфавиты с большим числом точек используются в ГКБП с большим отношением сигнал/шум или, что то же самое, с большими собственными значениями K_i .

Средняя мощность на выходе i -го ГКБП имеет вид

$$P_{bxi} = P_{qj} / |K_i|^2, \quad \mu_{j-1} \leq i \leq \mu_j - 1,$$

где $P_{qj} = \Delta^2 \varphi(2^{qj})$ — средняя мощность сигнала АФМ на входе и выходе ГКБП с номерами μ_{j-1} до $\mu_j - 1$, а

$$\varphi(q) = \left\{ \begin{array}{l} (2^q - 1/2)/6, q = 2m - 1, \\ (2^q - 1)/6, q = 2m, m = 1, 2, \dots \end{array} \right\}.$$

Средняя мощность на входе ГКМСИ-Ц ограничена величиной P_1

$$\frac{1}{N} \sum_{i=0}^{M_1-1} P_{bxi} \leq P_1.$$

Подставляя два последних выражения одно в другое, получаем

$$\Delta^2 \sum_{j=1}^Q \varphi(2^{qj}) \frac{1}{N} \sum_{i=\mu_{j-1}}^{M_j-1} \frac{1}{|K_i|^2} \leq P_1.$$

С учетом того, что

$$\frac{1}{N} \sum_{i=\mu_{j-1}}^{\mu_j-1} \frac{1}{|K_i|^2} = f_\mu(\mu_j) - f_\mu(\mu_{j-1}),$$

где $f_N(M) = (1/N) \sum_{i=0}^{M-1} 1/|K_i|^2$, получаем

$$\Delta^2 \sum_{j=1}^Q \varphi(2^{qj}) [f_N(\mu_j) - f_N(\mu_{j-1})] \leq P_1.$$

Максимальная скорость в каждом ГКБП при фиксированном q_j определяется следующим образом:

$$R(q_j, P_{qj}/\sigma_w^2) = R(q_j, \Delta^2 \varphi(2^{qj})/\sigma_w^2).$$

Суммарная скорость в ГКМСИ-Ц задается выражением

$$R = \varphi_0 \frac{1}{N} \sum_{j=1}^Q \nu_j R(q_j, \Delta^2 \varphi(2^{qj})/\sigma_w^2),$$

где $\nu_j = \mu_j - \mu_{j-1}$, $\mu_0 = 0$ — число ГКБП с одним и тем же алфавитом КАМ.

Оптимизация рассмотренного варианта по скорости при ограниченной средней мощности сигнала на входе канала свелась к выбору оптимального разбиения параллельных ГКБП на группы с одинаковой скоростью, оптимальному выбору алфавитов КАМ и минимального расстояния Δ в них. Это значит, что справедливо следующее утверждение.

Максимальная скорость, достижимая в ГКМСИ-Ц с предсказаниями и произвольными алфавитами КАМ в каждом из параллельных ГКБП при условии,

что минимальное расстояние во всех алфавитах постоянно и равно Δ , задается выражением

$$R_{\max} = \max_{\Delta > 0} \max_{\nu_j=1,2,\dots} \max_{q_j=1,2,\dots} \varphi_0 \frac{1}{N} \sum_{j=1}^Q \mu_j R(q_j, \frac{\Delta^2 \varphi(2^{qj})}{\sigma_w^2})$$

при ограничениях, приведенных выше, на допустимую среднюю мощность сигнала на входе ГКМСИ-Ц, а $\nu_j = \mu_j - \mu_{j-1}$, $\mu_0 = 0$, $0 < \mu_1 < \mu_2 < \dots < \mu_Q \leq M_1$ — разбиение множества ГКБП на группы из ν_j параллельных каналов, в каждом из которых используется один и тот же алфавит КАМ со средней мощностью

$$P_{qj} = \Delta^2 \varphi(2^{qj}).$$

Тем самым мы получили ступенчатую конструкцию, в которой передаваемый блок в результате виртуальных преобразований на передающей и приемной стороне преобразует систему к N параллельным каналам, из которых μ_1 модулируются КАМ2, μ_2 сигналов модулируются КАМ4, μ_3 сигналов модулируются КАМ8 и т. д. Последние μ_Q сигналов модулируются КАМ 2^Q .

Теперь на данную сигнальную конструкцию можно «накладывать» корректирующий код и получать СКК. Операцию наложения кода можно делать независимо на каждый ГКБП, т. е. во временной области. Однако гораздо более привлекательно выглядит построение СКК в частотной области, когда единым блоком СКК «накрываются» все каналы в одном временном блоке. Такой вариант оказывается наиболее предпочтительным, так как он практически не увеличивает задержки. В принципе возможно применение любой из блоковых или сверточных СКК, описанных выше. Необходима лишь достаточно простая модификация СКК, учитывающая разность алфавитов в последовательных символах. В результате получается так называемая ступенчатая СКК, кодируемая и декодируемая в пределах одного блока преобразования сигналов.

В данном разделе совершенно не затронуты вопросы оценки импульсной характеристики канала для точного выполнения OFDM-преобразований на передаче и приеме. Наиболее простой подход, используемый на практике, — отведение некоторого числа равномерно расставленных подканалов, используемых только для оценки характеристик канала. Особенно важно это в радиоканалах с меняющимися параметрами.

ГЛАВА 6

Беспроводные локальные сети стандарта IEEE 802.11 и оценка их производительности

6.1 Основные принципы IEEE 802.11

Рассмотрим подробнее, что представляет собой стандарт IEEE 802.11 [222] как базовый для всех последующих спецификаций. Как и все стандарты комитета IEEE 802, IEEE 802.11 рассматривает два нижних уровня модели взаимодействия открытых систем (OSI): физический и уровень звена данных Data Link layer (его же называют канальным уровнем). Причем последний подразделяется на два подуровня. Верхний — Logical Link Control (LLC) — описан в стандарте IEEE 802.2. Стандарт IEEE 802.11 рассматривает нижний подуровень — Medium Access Control (MAC), т. е. управление доступом к каналу (среде передачи). Иными словами, на физическом уровне стандарт определяет способ работы со средой передачи, скорость и методы модуляции. На MAC-уровне — принцип, по которому устройства используют (делят) общий канал, способы подключения устройств к точкам доступа и их аутентификации, механизмы защиты данных. Поскольку стандарт IEEE 802.11 разрабатывался как «беспроводной Ethernet», он предусматривает пакетную передачу с 48-битовыми адресами пакетов, как и любая сеть Ethernet. Комитет IEEE 802 особое внимание уделял совместимости всех своих стандартов, в результате проводные и беспроводные сети IEEE 802 легко сопрягаются друг с другом.

Когда речь заходит о радиотракте, ключевой вопрос — частотный диапазон. Как любой американский продукт, IEEE 802.11 привязан к действующим в США правилам. Изначально он был ориентирован на диапазон 2,400–2,4835 ГГц с шириной полосы 83,5 МГц. Определяемая стандартом спектральная маска для одного канала приведена на рис. 6.1 (мощность отсчитывается относительно пиков функции $\sin(x)/x$). Ширина канала по уровню -30 дБ составляет 22 МГц, следовательно, в полосе 83,5 МГц возможно три неперекрывающихся канала.

Стандарт предусматривает два основных способа организации локальной сети: по принципу «равный с равным» (ad-hoc-сеть — рис. 6.2, а) и в виде структу-

рированной сети (рис. 6.2, б). В первом случае связь устанавливается непосредственно между двумя станциями, и никакого администрирования не предусмотрено. В случае структурированных сетей (а как показала практика, это основной способ построения сетей IEEE 802.11) в их составе появляется дополнительное устройство — точка доступа (АР — Access Point), как правило, стационарная и действующая на фиксированном канале. Связь между устройствами происходит только через АР. Через них же возможен выход во внешние проводные сети. В сети IEEE 802.11 может быть несколько АР, объединенных проводной сетью Ethernet. Фактически такая сеть представляет собой набор базовых станций с перекрывающимися зонами охвата. Стандарт IEEE 802.11 допускает перемещения устройств из зоны одной АР в зону другой (роуминг), тем самым обеспечивая мобильность. Поскольку для мобильных станций важен вопрос ресурса элементов питания, в стандарте заложен специальный протокол управления энергопотреблением — непосредственно при обмене передающее устройство может перевести приемник в режим ожидания.

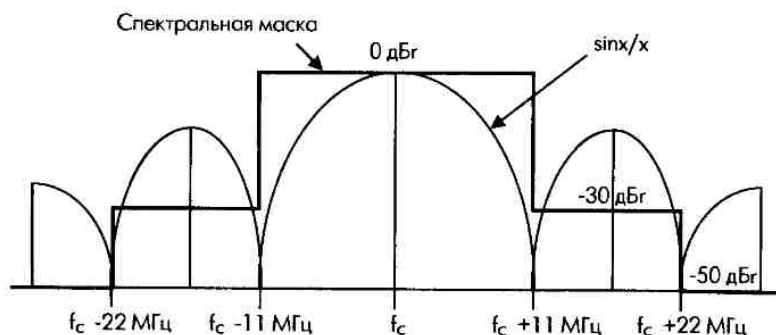


Рис. 6.1. Спектральная маска канала сети 802.11 при модуляции методом DSSS

Как правило, функции управления распределены между всеми устройствами сети IEEE 802.11 — так называемый режим распределенного управления DCF (Distributed coordination function). Однако для структурированных сетей возможен режим централизованного управления PCF (Point coordination function), когда управление передано одной определенной точке доступа. Необходимость в режиме PCF возникает при передаче чувствительной к задержкам информации (например, потоков видеoinформации). Ведь до настоящего времени сети семейства IEEE 802 действуют по принципу конкурентного доступа к каналу — приоритетов не существует (в ближайшее время ожидается появление версии IEEE 802.11e [176], определяющей распределенное управление с приоритетами). Чтобы их при необходимости задавать, и введен режим PCF. Однако работа в данном режиме может происходить только в определенные периодически повторяющиеся интервалы.

Важнейшее требование к стандартам беспроводной связи — это безопасность передачи данных. В связи с этим на MAC-уровне предусмотрен механизм защиты данных, включающий аутентификацию станций и собственно шифрование передаваемых данных. Этот механизм должен обеспечивать такой же уровень защиты, как и в обычных сетях Ethernet, поэтому его назвали WEP (Wired Equivalent

Privacy — эквивалент проводной конфиденциальности). Алгоритм WEP основан на использовании четырех общих для одной сети секретных ключей длиной 40 бит. Само шифрование происходит по алгоритму RC4 компании RSA Security. Алгоритм использует перемножение блоков исходных данных на псевдослучайную последовательность такой же длины, что и блок шифруемых данных. Генератор псевдослучайной последовательности инициализируется 64-разрядным числом, состоящим из 24-разрядного вектора инициализации (IV — initialization vector) и 40-разрядного секретного ключа. Существенно, что если секретный ключ известен устройствам сети и неизменен, то вектор IV может изменяться от пакета к пакету. Для защиты от несанкционированного изменения передаваемой информации каждый шифрованный пакет защищается 32-разрядной контрольной суммой (ICV — integrity check value). Таким образом, при шифровании к передаваемым данным добавляется 8 байт: 4 для ICV, 3 для IV, и еще 1 байт содержит информацию о номере используемого секретного ключа (одного из четырех). Отметим, что секретный ключ может быть гораздо длиннее — 64, 128 бит и т. д. Это не противоречит стандарту, более того, такое оборудование выпускается, однако законодательство США препятствует экспорту устройств, поддерживающих шифрование данных с ключом длиннее 40 бит. Именно поэтому производители ограничиваются 240 вариантами ключа.

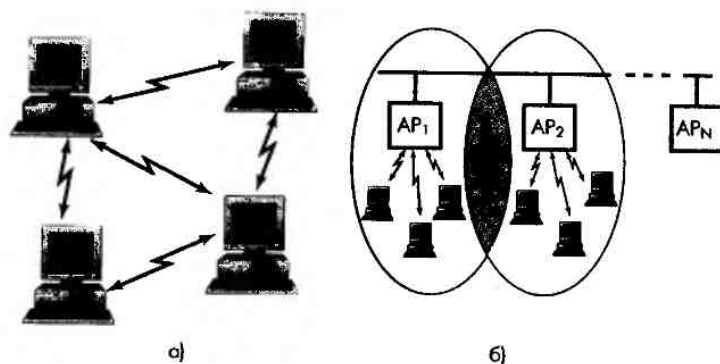


Рис. 6.2. Архитектура сети 802.11: а — ad-хос-сеть, б — структурированные сети

Защита информации в беспроводных сетях — это отдельная серьезная тема. Специальная группа работает над стандартом IEEE 802.11i, который также называют IEEE 802.X, поскольку закладываемые в него принципы применимы для различных сетей.

Отметим, что устройства, соответствующие исходной спецификации IEEE 802.11, практически не получили развития. Пропускная способность проводных сетей Ethernet сильно возросла, и максимальная скорость передачи по IEEE 802.11 — 2 Мбит/с — не удовлетворяла пользователей. Проблему на время решило появление стандарта IEEE 802.11b [161]. Благодаря расширенному диапазону скоростей (1; 2; 5,5; 11 Мбит/с) он завоевал массовую популярность. В этой спецификации описаны механизмы передачи в диапазоне 2,4 ГГц только посредством DSSS. Поэтому далее, без умаления общности, будем рассматривать IEEE 802.11 с точки зрения именно IEEE 802.11b.

6.2 МАС-уровень стандарта IEEE 802.11

Отличия стандарта IEEE 802.11 от других спецификаций семейства IEEE 802 начинаются на МАС-уровне. Как известно, основной принцип Ethernet — это множественный доступ к каналу связи с контролем несущей и обнаружением конфликтов (CSMA/CD — Carrier Sense Multiple Accseses with Collision Detection). Станция может начать передачу, только если канал свободен. Если станции обнаруживают, что на одном канале пытаются работать несколько станций, все они прекращают передачу и пытаются возобновить ее через случайный промежуток времени. Таким образом, даже при передаче устройство должно контролировать канал, т. е. работать на прием.

То, что относительно просто при проводной связи, проблематично в беспроводных коммуникациях — затухание сигнала в эфире намного сильнее, чем в проводе. Поэтому возникают две основные проблемы. Во-первых, весьма сложна, если вообще разрешима, задача контроля несущей передающим устройством (когда оно вещает, то собственный сигнал заведомо намного мощнее, чем сигнал удаленного устройства). Во-вторых, возможна ситуация, когда два устройства (А и В) удалены и не слышат друг друга, однако оба попадают в зону охвата третьего устройства С (рис. 6.3) — так называемая проблема скрытых станций. Если оба устройства А и В начнут передачу, то они принципиально не смогут обнаружить конфликтную ситуацию и определить, почему пакеты не проходят.

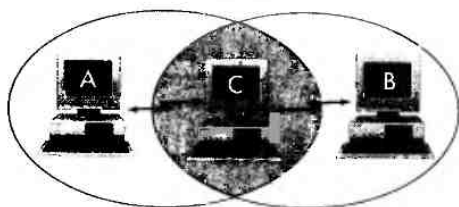


Рис. 6.3. Иллюстрация проблемы скрытых станций

Для устранения подобных проблем в спецификации IEEE 802.11 принят механизм CSMA/CA (Carrier Sense Multiple Accseses with Collision Avoidance) — множественный доступ с контролем несущей и предотвращением коллизий. Перед началом передачи устройство слушает эфир и дожидается, когда канал освободится. Канал считается свободным при условии, что не обнаружено активности в течение определенного промежутка времени — межкадрового интервала (IFS) определенного типа. Если в течение этого промежутка канал оставался свободным, устройство ожидает еще в течение случайного времени отсрочки и, если канал еще не занят, передает пакет. Если пакет предназначен конкретному устройству (не широковещательная или многоадресная передача), то приемник, успешно приняв пакет, посылает передатчику короткий кадр подтверждения получения ACK (ACKnowledge). Если передатчик не принял ACK, он считает посланный пакет утерянным и повторяет процедуру его передачи.

Примечательно, что, если устройство повторно передает пакет, для определения пезанятости канала оно должно использовать увеличенный межпакетный интервал (EIFS). Кроме того, время отсрочки выбирается случайным образом на некотором интервале. При первой попытке передачи этот интервал минимален. При каждой последующей он удваивается до тех пор, пока не достигнет заданного предельного значения. Эти меры приводят к тому, что устройство, успешно

передавшее пакет, имеет преимущества в захвате канала (кто ошибается, тот дольше ждет).

Перед первой попыткой получить доступ к каналу устройство загружает длительность случайного интервала отсрочки в специальный счетчик. Его значение декрементируется с заданной частотой, пока канал свободен. Как только счетчик обнулится, устройство может занимать канал. Если до обнуления счетчика канал занимает другое устройство, счет останавливается, сохраняя достигнутое значение. При следующей попытке отсчет начинается с сохраненной величины. В результате неуспевший в прошлый раз получает больше шансов занять канал в следующий. В проводных сетях Ethernet подобного механизма нет.

Однако описанные процедуры доступа не избавляют от проблемы скрытых станций. Для ее преодоления используются два дополнительных кадра: RTS (Request to Send — запрос на передачу) и CTS (Clear to Send — подтверждение готовности). Устройство, желающее отправить пакет-кадр данных, передает адресату короткий кадр RTS. Если приемное устройство готово к приему, оно выставляет передающему ответный кадр — CTS. Далее в соответствии с описанной выше процедурой передающее устройство отправляет кадр с данными и дожидается подтверждения ACK.

Подробнее работа данных механизмов MAC-уровня в локальных и региональных сетях описана соответственно в разделах 6.6 и 7.1.

Стандарт IEEE 802.11 предусматривает два механизма контроля за активностью в канале (обнаружения несущей): физический и виртуальный. Первый механизм реализован на физическом уровне и сводится к определению уровня сигнала в антенне и сравнению его с пороговой величиной. Виртуальный механизм обнаружения несущей основан на том, что в передаваемых кадрах данных, а также в управляющих кадрах ACK и RTS/CTS содержится информация о времени, необходимом для передачи пакета (или группы пакетов) и получения подтверждения. Все устройства сети получают информацию о текущей передаче и могут определить, сколько времени канал будет занят, т.е. устройство при установлении связи всем сообщает, на какое время оно резервирует канал.

Все описанные механизмы относятся к сети с распределенным управлением DCF. Однако в сети могут присутствовать и AP, наделенные полномочиями узурпировать управление, тогда их называют точками координации (PC). Когда сеть переходит в режим PCF, в трафике появляются интервалы, в которых конкурентный доступ отменен, и весь обмен происходит под управлением PC (рис. 6.4). По завершении такого интервала сеть возвращается в режим DCF. Интервалы под управлением PC следуют через строго определенный период, в начале каждого интервала PC выставляет особый сигнальный кадр (Beacon). PC не может передать очередной сигнальный кадр до тех пор, пока канал не освободится, т.е. очередной «свободный от конкуренции» интервал может начаться с задержкой.

Фактически режим PCF — это режим синхронной передачи, под который в асинхронной сети резервируются определенные интервалы. Этот режим позволяет использовать технологию IEEE 802.11 для таких приложений, как передача аудио/видео и других синхронных по своей природе данных.

Как мы уже говорили, весь обмен в сетях IEEE 802.11 происходит посредством отдельных кадров (frames). По их структуре особенно четко видно разделение на

физический и MAC-уровни. Фактически кадр формируется на MAC-уровне, на физическом уровне к нему добавляется заголовок физического уровня (PLCP). На MAC-уровень пакеты передаются от приложений верхнего уровня. Если их размер превышает максимально допустимый в IEEE 802.11, происходит дефрагментация — большой пакет разбивается на несколько меньших, которые передаются по специальной процедуре.

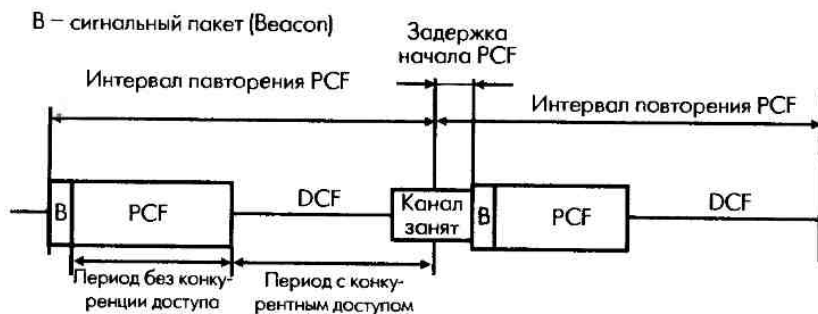


Рис. 6.4. Циклы работы сети в режимах с концентрированным (PCF) и распределенным (DCF) управлением

Кадры MAC-уровня могут быть трех типов: кадры данных, контрольные (ACK, RTS, CTS и т.п.) и кадры управления (например, Beacon). Их структура одинакова (рис. 6.5). Каждый кадр включает MAC-заголовок, информационное поле (Frame Body) и контрольную сумму CRC. В заголовке передается полная информация о версии протокола стандарта группы IEEE 802.11, типе кадра, системе защиты и т.д. (поле Frame Control); длительности процедуры передачи пакета (Duration/ID), адреса получателя/отправителя (Address1–4; четыре адресных поля необходимы, если пакеты передаются из подсети одной точки доступа в подсеть другой) и информация о последовательности связанных пакетов (Sequence Control). Информационное поле может быть различной длины или вообще отсутствовать (в контрольных кадрах).

6.3 Физический уровень стандарта IEEE 802.11b

На физическом уровне к MAC-кадрам (MPDU) добавляется заголовок физического уровня, состоящий из преамбулы и собственно PLCP-заголовка (см. рис. 6.5). Преамбула содержит стартовую синхропоследовательность (SYNC) для настройки приемника и 16-битный код начала кадра (SFD) — число F3A016. PLCP-заголовок включает поля SIGNAL (информация о скорости и типе модуляции), SERVICE (дополнительная информация, в том числе о применении высокоскоростных расширений и PBSS-модуляции) и LENGTH (время в микросекундах, необходимое для передачи следующей за заголовком части кадра). Все три поля заголовка защищены 16-битной контрольной суммой CRC.

В стандарте IEEE 802.11b предусмотрено два типа заголовков: длинный и короткий (рис. 6.6). Они отличаются длиной синхропоследовательности (128 и 56 бит), способом ее генерации, а также тем, что символ начала кадра в коротком

заголовке передается в обратном порядке. Кроме того, если все поля длинного заголовка передаются со скоростью 1 Мбит/с, то при коротком заголовке преамбула транслируется на скорости 1 Мбит/с, другие поля заголовка — со скоростью 2 Мбит/с. Остальную часть кадра можно передавать на любой из допустимых стандартом скоростей передачи, указанных в полях SIGNAL и SERVICE. Короткие заголовки физического уровня предусмотрены спецификацией IEEE 802.1b для увеличения пропускной способности сети.

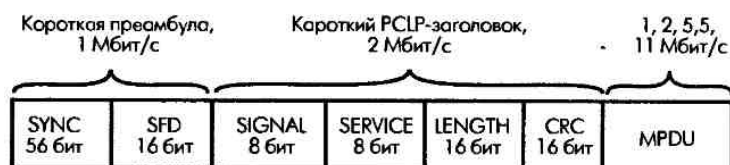


Рис. 6.5. Структура кадров сети 802.11 MAC-уровня

Из описания процедур связи сети IEEE 802.11 видно, что «накладные расходы» в этом стандарте выше, чем в проводной сети Ethernet. Поэтому крайне важно обеспечить высокую скорость передачи данных в канале. Повысить пропускную способность канала с заданной шириной полосы частот можно, разрабатывая и применяя более совершенные методы модуляции. По этому пути пошла группа разработчиков IEEE 802.11b.

Напомним, что изначально стандарт IEEE 802.11 предусматривал работу в режиме DSSS с использованием так называемой Баркеровской последовательности (Barker) длиной 11 бит: $B_1 = (10110111000)$. Каждый информационный бит замещается своим произведением по модулю 2 (операция «исключающее ИЛИ») с данной последовательностью, т. е. каждая информационная единица заменяется на B_1 , каждый ноль — на инверсию B_1 . В результате бит заменяется последовательностью 11 чипов. Далее сигнал кодируется посредством дифференциальной двух- или четырехпозиционной фазовой модуляции (DBPSK или DQPSK, один или два чипа на символ соответственно). При частоте модуляции несущей 11 МГц общая скорость составляет в зависимости от типа модуляции 1 и 2 Мбит/с.

Рис. 6.6. Короткий заголовок кадров сети 802.11b



Стандарт IEEE 802.11b дополнительно предусматривает скорости передачи 11 и 5,5 Мбит/с. Для этого используется так называемая ССК-модуляция (Complementary Code Keying — кодирование комплементарным кодом). В основу данного

метода легли работы специалистов компаний Intersil (ранее — Harris Semiconductor) и отчасти Agere Systems (полупроводниковое подразделение, выделенное из Lucent Technologies). Метод использует DQPSK-модуляцию в радиотракте. ССК-модуляция строится на выделении из последовательного информационного потока групп по 8 бит (d_0 – d_7). Эти 8 бит определяют информационный символ C из восьми комплексных чипов $C = [c_0, \dots, c_7]$. Чипы являются комплексными, поскольку они определяют I- и Q-квадратурные составляющие сигнала для DQPSK. Восемь чипов информационного символа последовательно модулируют несущую с частотой модуляции 11 МГц. Фактически, каждый чип представляет собой сигнал несущей частоты с фазовым сдвигом, определяемым формулой для всего символа

$$C = [c_0, \dots, c_7] = [j_1 + j_2 + j_3 + j_4; \quad j_1 + j_3 + j_4; \quad j_1 + j_2 + j_4; \\ -(j_1 + j_4); j_1 + j_2 + j_3; j_1 + j_3; -(j_1 + j_2); j_1]. \quad (6.1)$$

Элемент j_1 входит в каждый чип символа, т.е. изменяет фазу всего символа. Фазовая модуляция сигнала называется дифференциальной, поскольку значение j_1 текущего символа определяется относительно значения j_1 предшествовавшего символа. Изменение фазы j_1 задается как $D_{j_1} = (d_0 d_1) \pi / 2$ для четных символов, где $(d_0 d_1)$ — двухразрядное двоичное число (0–3). Для нечетных символов значение j_1 дополнительно смещают на π . Как мы уже отмечали, в стандарте IEEE 802.11b посредством ССК кодируют только MAC-кадр. Поэтому первый ССК-символ — тот, что следует сразу за символами заголовка физического уровня. Он имеет номер «0», т.е. является четным. Фаза последнего QPSK-символа заголовка кадра является опорной для определения j_1 первого символа информационного поля — к нему добавляется D_{j_1} . Остальные три параметра имеют абсолютные значения. Для скорости 11 Мбит/с они определяются как $j_2 = (d_2 d_3) \cdot \pi / 2$; $j_3 = (d_4 d_5) \cdot \pi / 2$; $j_4 = (d_6 d_7) \cdot \pi / 2$, где $(d_i d_{i+1})$ — двухразрядные двоичные числа.

Для скорости 5,5 Мбит/с используется DBPSK-модуляция, и ССК-символ определяют не восемь, а четыре информационных бита (d_0 – d_3), поэтому и скорость вдвое ниже. Параметр j_1 вычисляется так же, как и для 11 Мбит/с, остальные — иначе: $j_2 = d_2 \cdot \pi + \pi / 2$; $j_3 = 0$; $j_4 = d_3 \pi$.

В чем достоинство ССК-модуляции? Как видно из формулы (6.1), чипы символа (вектора) C определяются на основе последовательностей Уолша–Адамара. Для скорости 11 Мбит/с вектор входных данных $D = [d_0, \dots, d_7]$ можно представить в виде последовательности четверичных чисел $m_1 - m_4$, где $m_1 = (d_0 d_B)$, $m_2 = (d_2 d_3)$, $\dots$, $m_4 = (d_6 d_7)$. Пусть $M = [m_1, \dots, m_4]$. Тогда вектор $C = (M \times (WH + B)) \pi / 2$, где WH — верхняя половина матрицы Уолша–Адамара:

$$\{WH\} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Вектор $B = [0, 0, 0, 2, 0, 0, 2, 0]$, физически соответствующий повороту фазы чипов c_3 и c_6 на 180° (знаки «минус» в формуле (6.1)), необходим для улучшения корреляционных свойств кодовых последовательностей. Последовательности Уолша–Адамара хорошо изучены, обладают отличными автокорреляционными

свойствами. Что немаловажно, каждая такая последовательность мало коррелирует сама с собой при фазовом сдвиге — очень полезное свойство при борьбе с перетраженными сигналами. Нетрудно заметить, что теоретическое операционное усиление ССК-модуляции — 3 дБ (в два раза), поскольку без кодирования QPSK-модулированный с частотой 11 Мбит/с сигнал может транслировать 22 Мбит/с. Как видно, ССК-модуляция представляет собой вид блочного кода, а потому достаточно проста при аппаратной реализации. Совокупность этих свойств и обеспечила ССК место в стандарте IEEE 802.11b в качестве обязательного вида модуляции.

На практике важно не только операционное усиление. Существенную роль играет и равномерность распределения символов в фазовом пространстве — они должны как можно дальше отстоять друг от друга, чтобы минимизировать ошибки их детектирования. И с этой точки зрения ССК-модуляция не выглядит оптимальной, ее реальное операционное усиление не превышает 2 дБ. Поэтому изначально прорабатывался другой способ модуляции — пакетное бинарное сверточное кодирование PBCC (Packet Binary Convolutional Coding). Этот метод вошел в стандарт IEEE 802.11b как дополнительная (необязательная) опция. Механизм PBCC (рис. 6.7) был предложен специалистами фирмы Alantro Communications, в 2000 году вошедшей в состав компании Texas Instruments. PBCC позволяет добиваться в сетях IEEE 802.11b пропускной способности 5,5; 11 и 22 Мбит/с.



Рис. 6.7. Общая схема PBCC-модуляции

Как следует из названия, метод основан на сверточном кодировании. Для скоростей 5,5 и 11 Мбит/с поток информационных битов поступает в шести-разрядный сдвиговый регистр с сумматорами (рис. 6.8.а). В начальный момент времени все триггеры сдвигового регистра инициализируют нулем. В результате каждый исходный бит d заменяется двумя битами кодовой последовательности (c_0, c_1) . При скорости 11 Мбит/с c_0 и c_1 задают один символ четырехпозиционной QPSK-модуляции. Для скорости 5,5 Мбит/с используют двухпозиционную BPSK-модуляцию, последовательно передавая кодовые биты c_0 и c_1 . Если же нужна скорость 22 Мбит/с, схема кодирования усложняется (рис. 6.8.б): три кодовых бита (c_0-c_2) определяют один символ в 8-позиционной 8-PSK-модуляции.

После формирования PSK-символов происходит скремблирование. В зависимости от сигнала s (см. рис. 6.7) символ остается без изменений ($s = 0$), либо его фаза увеличивается на $\pi/2$ ($s = 1$). Значение s определяет 256-битовая циклически повторяющаяся последовательность S . Она формируется на основе начального вектора $U = 338B_{16}$, в котором равное число нулей и единиц. S представляет собой 16 последовательных векторов U_i ($i = 0, 15$), причем каждый вектор U_i циклически сдвинут влево по сравнению с U_{i-1} на $3i$ разряда.

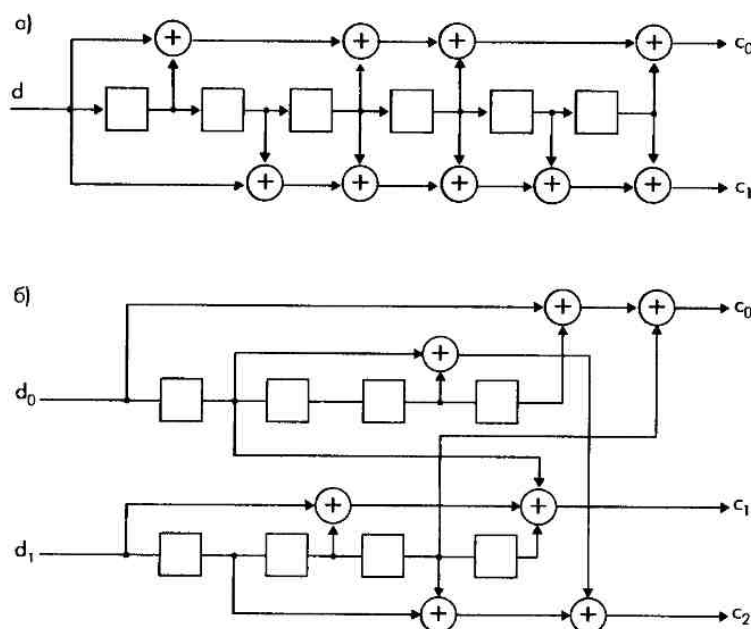


Рис. 6.8. Схема сверточного ВСС-кодирования: а — для скоростей 5,5 и 11 Мбит/с; б — для скорости 22 Мбит/с

У шестиразрядного сдвигового регистра, применяемого в РВСС для скоростей 11 и 5,5 Мбит/с, 64 возможных выходных состояния. Так что при модуляции РВСС информационные биты в фазовом пространстве оказываются гораздо дальше друг от друга, чем при ССК-модуляции. Поэтому РВСС и позволяет при одних и тех же соотношении сигнал/шум и уровне ошибок вести передачу с большей скоростью, чем в случае СКК (рис. 6.9). Однако плата за более эффективное кодирование — сложность аппаратной реализации данного алгоритма. Видимо, не случайно продвигать РВСС-модуляцию стала компания Texas Instruments — ведущий производитель процессоров цифровой обработки сигнала.

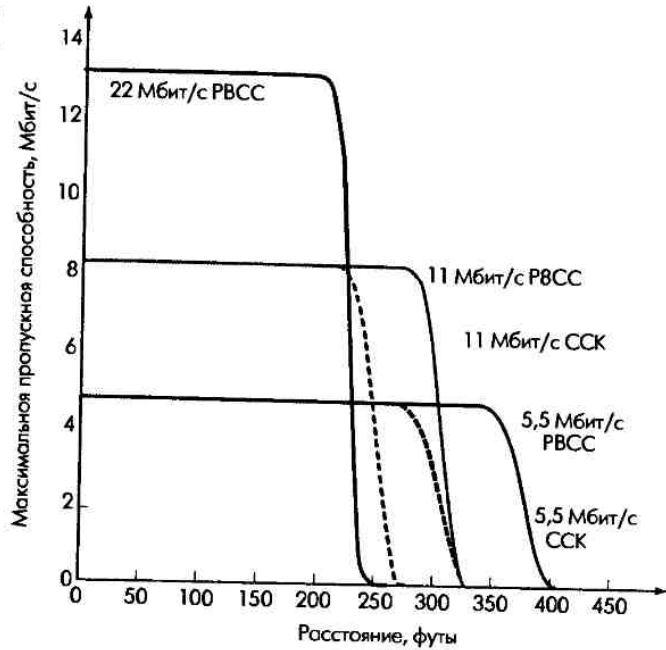
6.4 Аппаратная реализация

Чипсеты для оборудования сетей IEEE 802.11b производят многие компании. Кроме упомянутой Texas Instruments, прежде всего следует отметить линейку чипсетов PRISM корпорации Intersil. Первый ее чипсет для беспроводных ЛВС включал восемь основных микросхем и сразу привлек к себе внимание разработчиков аппаратуры. Выпущенный же более двух лет назад чипсет PRISM 2 можно рассматривать как базовый для последующих модификаций. В состав PRISM 2 (рис. 6.10) входят четыре ИС: MAC-контроллер HFA3841, baseband-процессор¹

¹Baseband-процессор — специализированный «коммуникационный» процессор, выполняющий все необходимые процедуры обработки данных на физическом уровне.

HFA3861B, квадратурный модулятор/демодулятор на промежуточной частоте (ПЧ) HFA3783 и ВЧ/ПЧ-конвертер HFA3683A. Как правило, для передачи требуется высокочастотный антенный усилитель, например HFA3963. Менее чем через год появилась модификация PRISM 2,5, в которой MAC-контроллер и baseband-процессор были интегрированы в одной ИС (ISL3873B), поддерживающей к тому же интерфейсы PCMCIA и USB.

Рис. 6.9. Дальность связи при различных способах модуляции и фиксированном затухании в канале (по материалам компании Texas Instruments о применении ИС ACX100)



Следующим шагом стало создание к лету 2001 года чипсета PRISM 3 на основе фирменной архитектуры прямого преобразования частоты ZIF (Zero IF, нулевая ПЧ) без использования ПЧ. В результате при цене чипсета 22 долл. стоимость всех компонентов для адаптера на его основе снижается, по мнению разработчиков, до 35 долл.

Отметим и объявленный чипсет компании Philips Semiconductors, первая ИС которого — однокристалльный трансивер SA2400 для диапазона 2,4 ГГц — была анонсирована в июле 2001 года. Данная ИС также построена по принципу прямого преобразования (без ПЧ) и содержит блоки приемника, передатчика, квадратурного модулятора/демодулятора, задающего генератора на основе ГУН и fractional-N-синтезатора, выходного малошумящего усилителя. В составе ИС — полностью интегрированные полосовые каналные фильтры, устройство автоматического контроля усиления, трехпроводная шина управления основными блоками устройства, интерфейс для сопряжения с baseband-процессором и т. д. Вторая микросхема чипсета — SA2440 — это интегрированные baseband-процессор и MAC-контроллер. Об аналогичном чипсете (Alchemy, который включает трансивер Am1770 и baseband/MAC-процессор Am1771) объявила и компания AMD.

В ноябре 2002 года корпорация Agere Systems анонсировала свой новый чипсет, в состав которого входят две ИС: контроллер физического уровня и МАС-контроллер. ИС WL1141 — однокристалльный контроллер физического уровня, включающий как baseband-процессор, так и аналоговую ВЧ-часть (рис. 6.11). Вместе с МАС-контроллером WL60010 он образует полный чипсет для устройств IEEE 802.11b.

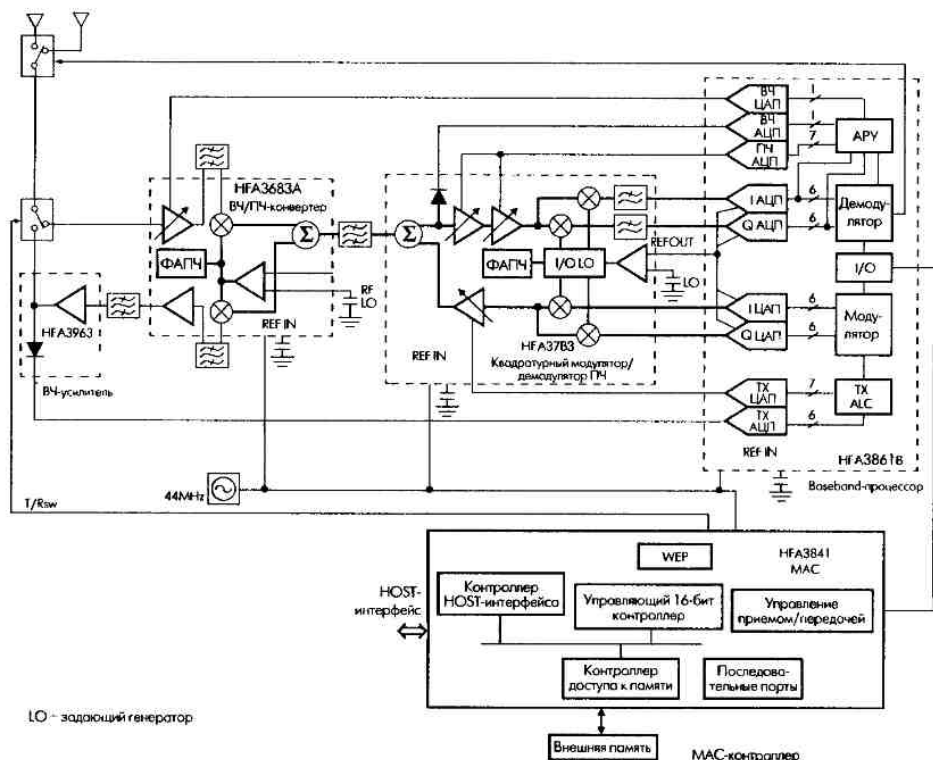


Рис. 6.10. Структурная схема устройства на базе чипсета PRISM 2

Надо отметить, что продукты компаний Intersil, Agere Systems, Philips поддерживают только ССК-модуляцию. PBCC реально присутствует практически только в продуктах компании Texas Instruments. Первой ИС, реализующей PBCC, стала ACX100 — однокристалльный МАС-контроллер с baseband-процессором (процессором телекоммуникационного обмена, т.е. выполняющим функции обработки, непосредственно связанные с каналом передачи). Эта ИС, кроме обязательных для стандарта IEEE 802.11b опций, поддерживает PBCC-модуляцию со скоростями 5,5; 11 и 22 Мбит/с. Среди других особенностей микросхемы — аппаратная поддержка 32-разрядных шин CardBus и PCI, последовательного интерфейса USB 1.1, а также интерфейса Ethernet. Интересно отметить, что при испытаниях устройств на основе этой схемы при одном и том же уровне пакетных ошибок ($PER = 10^{-2}$) и скорости передачи 11 Мбит/с работа в режиме ССК была возможной при отношении сигнал/шум 8,5 дБ, а применение PBCC позволяло ухудшить это соотношение до 4,5 дБ.

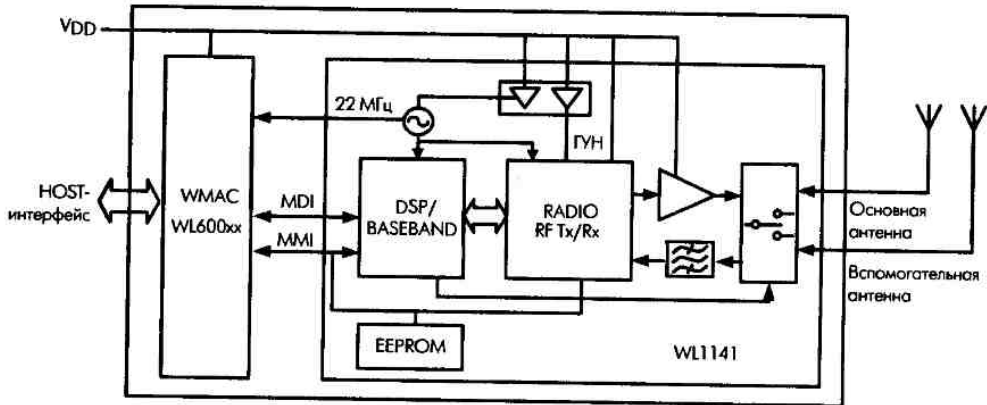


Рис. 6.11. Чипсет компании Agere Systems

Однако, несмотря на все усилия Texas Instruments, режим 22 Мбит/с так и не вошел в стандарт IEEE 802.11b, хотя даже появилось обозначение «IEEE 802.11b+» и устройства на основе ACX100 начали производить такие известные компании, как D-Link и NDC. Не получили распространения и сети со скоростью 22 Мбит/с. Видимо, тут играют роль два обстоятельства. Прежде всего, процедура стандартизации — это конкуренция за получение значительных финансовых прибылей в виде лицензионных отчислений тому, чья технология стала стандартной. В данном случае столкнулись интересы двух гигантов полупроводниковой индустрии — компаний Intersil и Texas Instruments. Поддерживать одновременно два вида модуляции сложно в финансовом и техническом отношении, поэтому и был принят ССК-вариант.

С другой стороны, скорости порядка 22 Мбит/с стали мало интересны для потребителей, так как перед ними открылись новые перспективы — до 54 Мбит/с.

6.5 Стандарты IEEE 802.11a и 802.11g

Стандарт IEEE 802.11a появился практически одновременно с IEEE 802.11b, в сентябре 1999 года. Эта спецификация была ориентирована на работу в диапазоне 5 ГГц и основана на принципиально ином, чем описано выше, механизме кодирования данных — на частотном мультиплексировании посредством ортогональных несущих (OFDM). Как мы уже отмечали, в IEEE 802.11a каждый кадр передается посредством 52 ортогональных несущих, каждая с шириной полосы порядка 300 кГц (20 МГц/64). Ширина одного канала — 20 МГц. Несущие модулируются посредством BPSK, QPSK, а также 16- и 64-позиционной квадратурной амплитудной модуляции (QAM). В совокупности с различными скоростями кодирования r ($1/2$ и $3/4$, для 64-QAM — $2/3$ и $3/4$) образуется набор скоростей передачи 6, 9, 12, 18, 24, 36, 48 и 54 Мбит/с.

Из 52 несущих 48 предназначены для передачи информационных символов, остальные 4 — служебные. Структура заголовков физического уровня отличается от принятого в спецификации IEEE 802.11b, но не существенно (рис. 6.12). Кадр

включает преамбулу (12 символов синхропоследовательности), заголовок физического уровня (PLCP-заголовок) и собственно информационное поле, сформированное на MAC-уровне. В заголовке передается информация о скорости кодирования, типе модуляции и длине кадра. Преамбула и заголовок транслируются с минимально возможной скоростью (BPSK, скорость кодирования $r = 1/2$), а информационное поле — с указанной в заголовке, как правило максимальной, скоростью, в зависимости от условий обмена. OFDM-символы передаются через каждые 4 мкс, причем каждому символу длительностью 3,2 мкс предшествует защитный интервал 0,8 мкс (повторяющаяся часть символа). Последний необходим для борьбы с многолучевым распространением сигнала — отраженный и пришедший с задержкой символ попадет в защитный интервал и не повредит следующий символ.

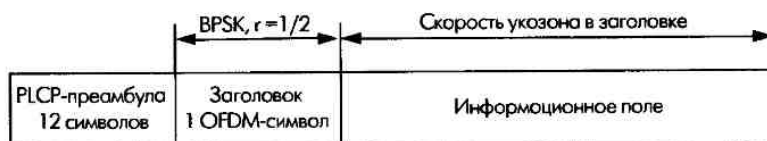


Рис. 6.12. Структура заголовка физического уровня стандарта IEEE 802.11a

Естественно, формирование/декодирование OFDM-символов происходит посредством быстрого преобразования Фурье (обратного/прямого, ОБПФ/БПФ). Функциональная схема трактов приема/передачи (рис. 6.13) достаточно стандартна для данного метода и включает сверточный кодер, механизм перемежения /перераспределения (защита от пакетных ошибок) и процессор ОБПФ. Фурье-процессор, собственно, и формирует суммарный сигнал, после чего к символу добавляется защитный интервал, окончательно формируется OFDM-символ и посредством квадратурного модулятора/конвертера переносится в заданную частотную область. При приеме все происходит в обратном порядке.

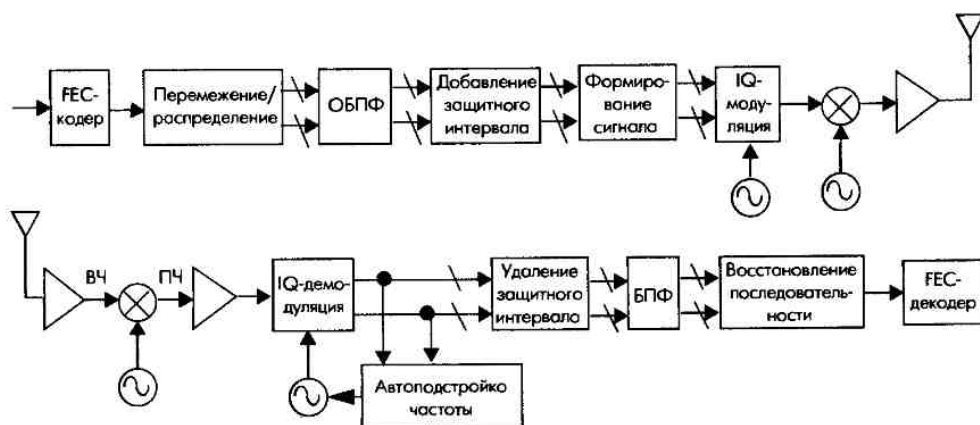


Рис. 6.13. Функциональная схема трактов приема/передачи стандарта IEEE 802.11a

Стандарт IEEE 802.11g по сути представляет собой перенесение схемы модуляции OFDM, прекрасно зарекомендовавшей себя в 802.11a, из диапазона 5 ГГц в

область 2,4 ГГц при сохранении функциональности устройств стандарта 802.11b. Это возможно, поскольку в стандартах 802.11 ширина одного канала в диапазонах 2,4 и 5 ГГц схожа — 22 МГц по уровню -30 и -20 дБ, соответственно. По уровню -28 дБ маска канала в 802.11a допускает спектральную полосу шириной 40 МГц.

Одним из основных требований к спецификации 802.11g была обратная совместимость с устройствами 802.11b. Это требование привело к очередному столкновению интересов компаний Intersil и TI. Действительно, в стандарте 802.11b в качестве основного способа модуляции принята схема ССК (Complementary Code Keying), а в качестве дополнительной возможности допускается модуляция PBSS. В последней крайне заинтересована компания TI. Разработчики 802.11g предусмотрели ССК-модуляцию для скоростей до 11 Мбит/с и OFDM для более высоких скоростей. С этим были согласны все. Но сети стандарта 802.11 при работе используют принцип CSMA/CA — множественный доступ к каналу связи с контролем несущей и предотвращением коллизий. Ни одно устройство 802.11 не должно начинать передачу, пока не убедится, что эфир в его диапазоне свободен от других устройств. Если в зоне слышимости окажутся устройства 802.11b и 802.11g, причем обмен будет происходить между устройствами 802.11g посредством OFDM, то оборудование 802.11b просто не поймет, что другие устройства сети ведут передачу, и попытается начать трансляцию. Последствия очевидны.

Чтобы подобную ситуацию не допустить, предусмотрена возможность работы в смешанном режиме — ССК-OFDM. Информация в сетях 802.11 передается кадрами. Каждый информационный кадр включает два основных поля: преамбулу с заголовком и информационное поле (рис. 6.14). Преамбула содержит синхропоследовательность и код начала кадра, заголовок — служебную информацию, в том числе о типе модуляции, скорости и продолжительности передачи кадра. В режиме ССК-OFDM преамбула и заголовок модулируются методом ССК (реально — путем прямого расширения спектра DSSS посредством последовательности Баркера, поэтому в стандарте 802.11g этот режим именуется DSSS-OFDM), а информационное поле — методом OFDM. Таким образом, все устройства 802.11b, постоянно «прослушивающие» эфир, принимают заголовки кадров и узнают, сколько времени будет транслироваться кадр 802.11g. В этот период они «молчат». Естественно, пропускная способность сети падает, поскольку скорость передачи преамбулы и заголовка — 1 Мбит/с.

Видимо, данный подход не устраивал лагерь сторонников технологии PBSS, и для достижения компромисса в стандарт 802.11g в качестве дополнительной возможности ввели, так же как и в 802.11b, необязательный режим — PBSS, в котором заголовок и преамбула передаются так же, как и при ССК, а информационное поле модулируется по схеме PBSS и передается на скорости 22



Рис. 6.14. Кадры IEEE 802.11g в различных режимах модуляции

или 33 Мбит/с. В результате устройства стандарта 802.11g должны оказаться совместимыми со всеми модификациями оборудования 802.11b и не создавать взаимных помех. Диапазон поддерживаемых им скоростей отражен в табл. 6.1, зависимость скорости от типа модуляции — на рис. 6.15.

Таблица 6.1. Возможные скорости и тип модуляции в спецификации IEEE 802.11g

Скорость, Мбит/с	Тип модуляции	
	Обязательно	Допустимо
1	Последовательность Баркера	
2	Последовательность Баркера	
5,5	ССК	PBCC
6	OFDM	ССК-OFDM
9		OFDM, ССК-OFDM
11	ССК	PBCC
12	OFDM	ССК-OFDM
18		OFDM, ССК-OFDM
22		PBCC
24	OFDM	ССК-OFDM
33		PBCC
36		OFDM, ССК-OFDM
48		OFDM, ССК-OFDM
54		OFDM, ССК-OFDM

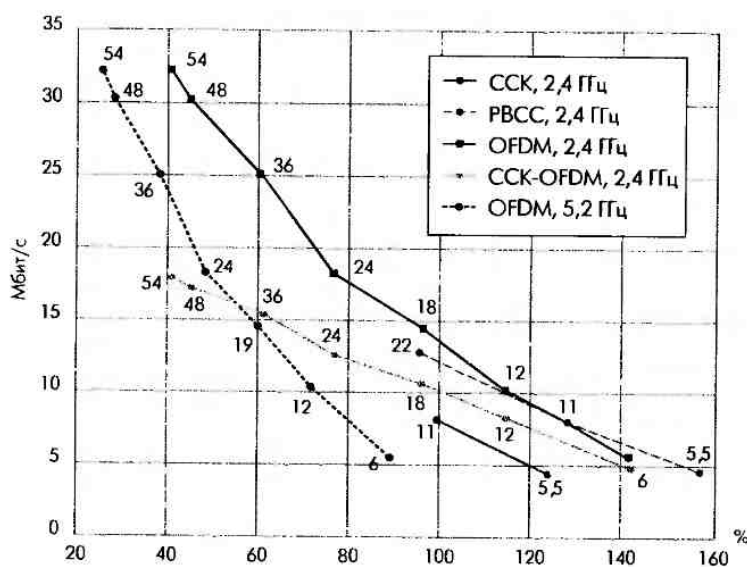


Рис. 6.15. Зависимость скорости передачи от расстояния для различных технологий передачи. Расстояние приведено в процентах, 100% — дальность передачи с модуляцией ССК на скорости 11 Мбит/с

Очевидно, что устройствам стандарта IEEE 802.11g достаточно долго придется работать в одних сетях с оборудованием 802.11b. Также очевидно, что производители в массе своей не будут поддерживать режимы CCK-OFDM и PBSS в силу их необязательности, ведь почти все решает цена устройства. Поэтому одна из основных проблем нового стандарта — как обеспечить бесконфликтную работу смешанных сетей 802.11b/g.

Основной принцип работы в сетях 802.11 — «слушать, прежде чем вещать». Но устройства 802.11b не способны услышать устройства 802.11g в OFDM-режиме. Ситуация аналогична проблеме скрытых станций: два устройства удалены настолько, что не слышат друг друга и пытаются обратиться к третьему, которое находится в зоне слышимости обоих. Для предотвращения конфликтов в подобной ситуации в 802.11 введен защитный механизм, предусматривающий перед началом информационного обмена передачу короткого кадра «запрос на передачу» (RTS) и получение кадра подтверждения «можно передавать» (CTS). Механизм RTS/CTS применим и к смешанным сетям 802.11b/g. Естественно, эти кадры должны транслироваться в режиме CCK, который обязаны понимать все устройства. Однако защитный механизм существенно снижает пропускную способность сети. Так, при физической скорости 54 Мбит/с потолок пропускной способности гомогенной сети 802.11g (с учетом всей служебной и управляющей информации) около 32 Мбит/с, а реальные показатели оборудования — на уровне 24 Мбит/с. Если же сеть смешанная, то защитный механизм RTS/CTS понизит пропускную способность до 12 Мбит/с. Это практически вдвое превышает пропускную способность однородной сети 802.11b (~6 Мбит/с), но ведь всегда хочется большего. Поэтому вместо механизма RTS/CTS можно использовать только кадры CTS, предшествующие каждому OFDM-кадру. В результате пропускная способность несколько повысится — до 14,5 Мбит/с. Однако этот механизм неприемлем, если не все устройства сети находятся в зоне слышимости друг друга (пресловутая проблема «скрытой точки»).

Видимо, поэтому производители ИС для сетей 802.11 разрабатывают специальные механизмы, способные в рамках действующих стандартов повысить скорость передачи. Так, компания Atheros для стандартов 802.11a и g предложила так называемый режим Turbo Mode, позволяющий удвоить номинальную скорость до 108 Мбит/с за счет передачи информации одновременно по двум каналам. Для поддержки Turbo Mode компания выпустила специальный чипсет AR5001X+, отличающийся от AR5001X модифицированным процессором AR5212.

Корпорация Intersil пошла другим путем. В апреле она представила свою технологию PRISM Nitro, включающую два основных элемента: защитный механизм и групповую передачу OFDM-кадров [217]. Защитный механизм не содержит ничего принципиально нового и подразумевает передачу перед каждым OFDM-кадром кадра CTS. Intersil ратует за введение этого средства защиты в спецификацию 802.11g в качестве обязательного элемента. Групповая же передача OFDM-кадров способна, по мнению специалистов компании, существенно повысить пропускную способность как смешанной 802.11b/g сети, так и однородной.

В случае смешанной сети предлагается каждому устройству предоставлять в трафике примерно равный временной интервал. Действительно, для передачи

ССК-кадра со скоростью 11 Мбит/с требуется примерно столько же времени, сколько для передачи шести OFDM-кадров со скоростью 54 Мбит/с (с учетом всех накладных расходов). Если устройства 802.11b и g поочередно передают одинаковый объем информации, на передачу, например, пакетов с информационным полем 1500 байт двум устройствам потребуется 2143 мкс (рис. 6.16).

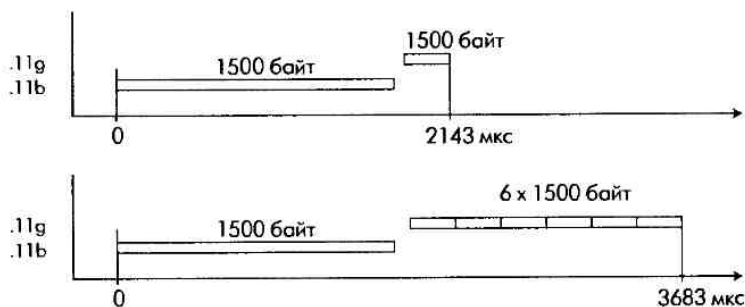


Рис. 6.16. Групповая передача OFDM-пакетов по технологии PRISM Nitro в смешанной сети

Если же каждому устройству для трансляции выделить равные временные интервалы, устройство 802.11g передаст шесть пакетов (9000 байт) — всего 10 500 байт за 3683 мкс. В первом случае пропускная способность сети составит 11,2 Мбит/с, во втором — 22,8 Мбит/с: выигрыш — более чем в два раза. В случае однородной 802.11g-сети групповая передача пакетов также дает выигрыш за счет того, что внутри группы между пакетами не требуется выставить кадр CTS и выждать межкадровый интервал. Необходим только короткий кадр подтверждения приема ACK (рис. 6.17).

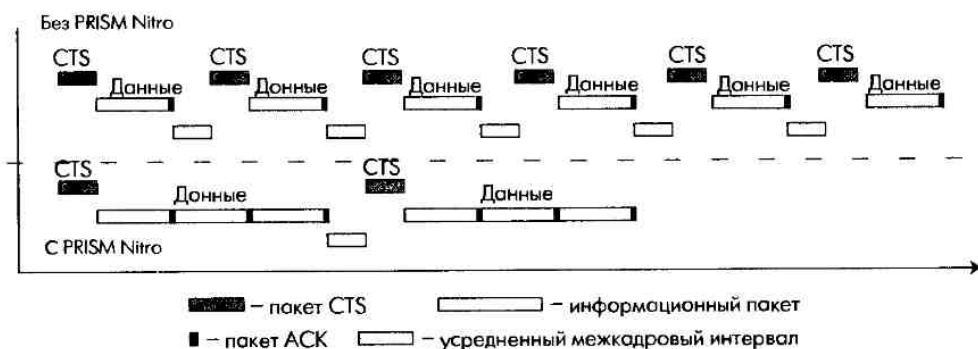


Рис. 6.17. Групповая передача OFDM-пакетов по технологии PRISM Nitro в однородной сети

Технология PRISM Nitro реализуется на уровне системного программного обеспечения. Она специально разрабатывалась для применения совместно с чип-сетями PRISM Duette и GT. Ее создатели утверждают, что она полностью соответствует требованиям спецификации IEEE 802.11g. Данная технология уже используется в новом маршрутизаторе CONNECT2AIR и сетевых картах ком-

пании Fujitsu Siemens Computers. Отметим, что эта компания первой применила новейший процессор фирмы Intersil для точек доступа/маршрутизаторов беспроводных сетей ISL3893. Этот процессор, построенный на базе микропроцессорного ядра ARM9, реализует функции как сетевого процессора беспроводной сети, так и сети Ethernet. Он предназначен для работы совместно с чипсетами PRISM GT и PRISM Duette.

6.6 Схема распределенного управления в локальных сетях

В режиме распределенного управления (схема DCF) информационные пакеты передаются в общем случае двумя способами. Короткие пакеты, чья длина не превышает некоторого предела $\bar{P}$, передаются с помощью механизма базового доступа. При этом механизме, показанном на рис. 6.18, станция, успешно принявшая кадр DATA, содержащий информационный пакет, спустя короткий интервал SIFS немедленно отвечает положительным подтверждением ACK.

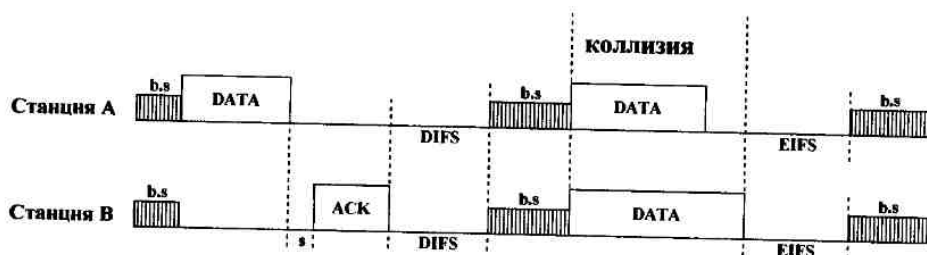


Рис. 6.18. Механизм базового доступа (s — SIFS, $b.s$ — слоты времени отсрочки)

Для пакетов с длиной, большей RTS-порога $\bar{P}$, используется механизм Request-To-Send/Clear-To-Send (RTS/CTS). В этом случае, изображенном на рис. 6.19, передача кадра DATA предваряется запросом на передачу (RTS), направляемым к принимающей станции, которая спустя SIFS отвечает разрешающим кадром CTS. И только по получении кадра CTS передается кадр DATA, который подтверждается кадром ACK. Таким образом, значение $\bar{P}$ выбирается в результате разумного компромисса между накладными расходами механизма RTS/CTS, заключающимися в передаче двух дополнительных кадров RTS и CTS, и выигрышем в длительности коллизии. Как видно из сравнения рис. 6.18 и 6.19, длительность коллизии при базовом доступе определяется максимальной длиной кадров DATA, участвующих в коллизии, а при использовании механизма RTS/CTS она равна длительности передачи короткого кадра RTS.

После завершения попытки передачи очередного пакета станция переходит в состояние отсрочки спустя интервал DIFS, если попытка была успешной (т. е. коллизия отсутствовала и все кадры, относящиеся к данному пакету, были переданы корректно, без искажения помехами), или EIFS — при неудачной попытке. При этом счетчик времени нахождения в состоянии отсрочки устанавливается в начальное значение b , которое называется временем отсрочки, измеряется в слотах

длительностью σ и равновероятно выбирается из множества $(0, \dots, w-1)$. Значение w , называемое конкурентным окном, зависит от значения n_r числа сделанных попыток передачи текущего пакета:

$$w = W_i = W_0 2^{n_r} \text{ при } n_r < m \text{ и } w = W_m \text{ при } n_r \geq m, \quad (6.2)$$

где $W_0 = CW_{\min}$ и $W_m = W_0 2^m = CW_{\max}$ — соответственно минимальное и максимальное конкурентные окна.

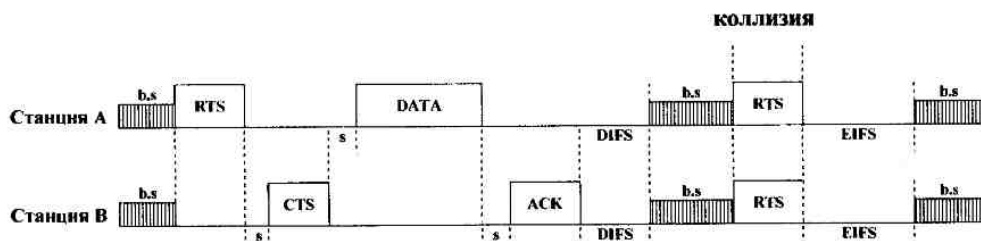


Рис. 6.19. Механизм RTS/CTS

В процессе передачи текущего пакета каждая станция, инициировавшая передачу, подсчитывает число неудачных попыток передачи коротких (n_s) и длинных (n_d) кадров. Пусть станция передала кадр DATA с пакетом, длина которого меньше или равна $\bar{P}$, или кадр RTS. Тогда, если в течение тайм-аута, равного EIFS, поступит ответный корректный кадр, соответственно ACK или CTS, то значение счетчика n_s обнуляется, а иначе значение n_s увеличивается на единицу. Аналогично значение счетчика n_d обнуляется или увеличивается на единицу в случаях соответственно приема или отсутствия (в течение тайм-аута) корректного кадра ACK, подтверждающего успешную передачу кадра DATA с пакетом, длина которого больше $\bar{P}$. Если значение любого из счетчиков n_s и n_d достигает некоторых предельных значений соответственно N_s и N_d , то текущий пакет отбрасывается и (при непустой очереди) выбирается следующий пакет для передачи с обнулением значений n_r , n_s и n_d .

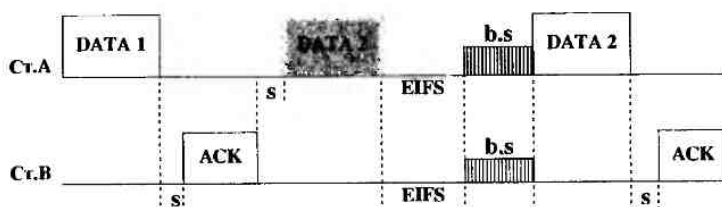


Рис. 6.20. Передача фрагментированных пакетов

Для снижения влияния помех стандарт [222] рекомендует разбивать пакеты, размер которых больше некоторого порога фрагментации L_f , на фрагменты, размер каждого из которых (кроме последнего) равен L_f . Таким образом, пакет передается в виде цепочки кадров DATA, содержащих последовательные фрагменты и перемежаемых ответными кадрами ACK, а также короткими межкадровыми промежутками SIFS (см. рис. 6.20). При искажении некоторого фрагмента

станция переходит в состояние отсрочки, увеличивая на единицу значение счетчика попыток n_r , и повторная передача начнется именно с этого фрагмента, а не с начала всего пакета.

Процесс передачи первого фрагмента в общем случае складывается из обмена четырьмя кадрами: запрос на передачу (кадр RTS), разрешение на передачу (кадр CTS), сам фрагмент с заголовком (кадр DATA) и подтверждение успешной передачи (кадр ACK). Эти фазы обмена разделяются коротким временным интервалом SIFS. При длине фрагмента, меньшей некоторого предела $\bar{P}$, кадры RTS и CTS не используются. Если станция в течение тайм-аута не получает ответные корректные кадры CTS и ACK (а также сразу по приеме искаженных кадров или кадров, относящихся к другой станции), она считает, что либо произошла коллизия, либо последний переданный кадр был искажен, и переходит в состояние отсрочки, увеличивая на единицу свой счетчик коллизий n_r . После успешной передачи фрагмента (по получении ACK) станция спустя SIFS начинает передачу следующего фрагмента. Передача последующих фрагментов пакета отличается лишь тем, что при первой попытке не используются кадры RTS и CTS независимо от длины фрагмента. После успешной передачи последнего фрагмента пакета станция переходит в состояние отсрочки, обнуляя счетчики n_r , n_s и n_d .

Станция начинает передачу при выполнении следующих условий: 1) истекло время отсрочки с момента последней передачи данной станции; 2) эфир этой станции был свободен в течение интервала задержки (равного DIFS, если последний «услышанный» кадр не был искаженным, а иначе — EIFS); 3) в очереди на передачу имеются готовые пакеты. В частности, при поступлении нового пакета в пустую очередь на передачу станция либо сразу начинает передачу кадра RTS или DATA (см. выше), если на момент поступления пакета эфир был свободен в течение интервала задержки, либо переходит в состояние отсрочки с $n_r = 0$. Этот переход в общем случае состоит из двух фаз: 1) ожидание освобождения канала, если он занят, и 2) интервал задержки. Если в течение этого интервала канал был свободен, станция начинает отсчет времени отсрочки (а иначе обе фазы перехода повторяются). Отсчет прекращается по получении сигнала о занятости канала и возобновляется только спустя интервал задержки с момента освобождения канала. По окончании отсчета времени отсрочки станция либо сразу же посылает кадр RTS (или DATA при длине фрагмента, меньшей $\bar{P}$), либо простаивает в ожидании нового пакета (при пустой очереди на передачу).

Для повышения надежности слежения за состоянием эфира протокол Radio-Ethernet наряду с регистрацией «физической» занятости эфира предусматривает механизм отслеживания «виртуальной» занятости. Для этого кадры содержат поле предполагаемой длительности передачи t_{nav} :

- для кадра RTS значение t_{nav} равно сумме времен передачи CTS, DATA и ACK;
- для CTS оно равно сумме времен передачи DATA и ACK;
- для DATA оно равно сумме времен передачи ACK данного фрагмента и DATA и ACK следующего фрагмента (если он есть);
- наконец, для ACK t_{nav} равно сумме времен передачи DATA и ACK следующего фрагмента.

Все эти суммы включают соответствующее число интервалов SIFS между кадрами. Станция, принявшая кадр, не предназначенный ей, считает из него значение t_{nav} и считает эфир «виртуально» занятым в течение соответствующего интервала (в случае кадра RTS механизм более сложен, см. [222]). Механизм отслеживания «виртуальной» занятости особенно полезен в случае скрытых станций, причем особенности схемы DCF в этом случае будут рассмотрены в следующей главе.

Изложение особенностей протокола для технологии FHSS также дадим в следующей главе, а сейчас перейдем к описанию используемых моделей БЛС на основе протокола IEEE 802.11.

6.7 Обобщение существующих методов оценки производительности беспроводных локальных сетей

В ранних работах пропускная способность 802.11 WLAN под управлением механизма DCF изучалась либо экспериментально (например, [219]), либо путем упрощенного моделирования [142, 162], основанного на предположениях, существенно упрощающих реальный алгоритм отсрочки. Наиболее детальный анализ схемы DCF приведен в работах [130, 137], в которых разработаны аналитические методы оценки пропускной способности беспроводной локальной сети (БЛС) 802.11 при высокой нагрузке, когда ко всем станциям БЛС всегда имеются непустые очереди (условие насыщения сети). Этот показатель, определенный в [130] как *saturation throughput* (пропускная способность в условиях насыщения), получен в предположении идеального канала передачи, т. е. при отсутствии шумов и скрытых станций.

Рассмотрим БЛС с протоколом IEEE 802.11 DCF, состоящую из N станций и работающую в условиях насыщения и идеального канала. Очевидно, при отсутствии помех число переповторов передачи пакета ограничено $R = N_s$. Здесь и далее в этой главе все станции БЛС считаются статистически однородными. Статистическая однородность станций заключается в одинаковом вероятностном распределении длин пакетов, выбираемых каждой станцией из очереди.

Следуя подходу, предложенному Бьянки в [130], применим следующую дискретную и целочисленную временную шкалу работы сети: моменты t и $t + 1$ соответствуют началам следующих друг за другом виртуальных слотов. Виртуальные слоты не одинаковы, и каждый из них может представлять из себя: 1) «пустой» слот отсрочки σ , в котором ни одна из станций не ведет передачу; 2) «успешный» слот, в котором только одна из станций передает пакет; 3) «коллизийный» слот, в котором две станции или более пытаются осуществить передачу.

Согласно модели Бьянки [130] предполагается, что в начале каждого слота каждая станция пытается отправить пакет с одинаковой вероятностью, равной τ . В модели Кали и др. [137] предполагается, что время отсрочки b не зависит от числа переповторов n_r текущей передачи и выбирается из геометрического распределения с параметром τ , т. е. $b = 0, 1, 2, \dots$ с вероятностями

$\tau, \tau(1 - \tau), \tau(1 - \tau)^2, \dots$ Очевидно, что данные предположения эквивалентны, если не учитывать эффект захвата канала, описываемый в следующем разделе, т.е. в обеих рассматриваемых моделях считается, что любая станция может начать передачу в начале каждого слота. Приняв любое из введенных предположений, нетрудно получить выражения для вероятностей «пустого» слота (p_e), «успешного» слота (p_s) и «коллизийного» слота (p_c):

$$p_e = (1 - \tau)^N, \quad p_s = N\tau(1 - \tau)^{N-1}, \quad p_c = 1 - (1 - \tau)^N - N\tau(1 - \tau)^{N-1}. \quad (6.3)$$

В данном случае вероятность коллизии не зависит от размера передаваемого пакета. Поэтому предполагается, что длительности t_s и t_c «успешного» и «коллизийного» слотов, в свою очередь, независимы от числа попыток (или нескольких значений числа попыток в случае коллизии) и определяются только функцией распределения вероятности $F(\cdot)$ для размера пакета данных $P \in [0, P_{\max}]$, нормализованного скоростью канала V_c . Исходя из рис. 6.18 и 6.19, находим средние значения t_s и t_c :

$$T_s = [1 - F(P_l)](t_{\text{RTS}} + t_{\text{CTS}} + 2\delta + 2 \text{ SIFS}) + H + E[P] + t_{\text{ACK}} + 2\delta + \text{SIFS} + \text{DIFS} \quad (6.4)$$

и

$$T_c = E[P^*] + \delta + \text{EIFS}, \quad (6.5)$$

где P_l является порогом RTS/CTS $\bar{P}$, нормализованным V_c ; t_{RTS} , $t_{\text{CTS}} = t_{\text{ACK}}$ и H — это времена, необходимые для передачи соответственно кадров RTS, CTS и ACK, а также заголовка DATA; δ — это задержка передачи, принятая одинаковой для всех пар станций; $E[P]$ — среднее значение P ; $E[P^*]$ — средняя продолжительность коллизии, т.е. среднее время, требуемое для передачи самого длинного кадра, вовлеченного в коллизию.

Последняя величина, очевидно, определяется по формуле

$$E[P^*] = p_c^{-1} \sum_{k=2}^N \left(\frac{N}{k} \right) \tau^k (1 - \tau)^{N-k} E[P^*|k], \quad (6.6)$$

где $E[P^*|k]$ — это средняя продолжительность коллизии, в которой участвуют k станций. Чтобы найти $E[P^*|k]$, мы примем предположения Бьянки [130]: а) функция распределения $F(\cdot)$ одинакова для всех станций; б) существует производная $f(P) = dF/dP$ для всех $P \in (0, P_{\max})$. Тогда

$$\begin{aligned} E[P^*|k] &= t_{\text{RTS}}[1 - F(P_l)]^k + k \int_0^{P_l} (x + H)f(x)[F(x) + 1 - F(P_l)]^{k-1} dx = \\ &= H + P_l - (H - t_{\text{RTS}})[1 - F(P_l)]^k - \int_0^{P_l} [F(x) + 1 - F(P_l)]^k dx. \end{aligned} \quad (6.7)$$

Подставив (6.7) в (6.6), после простых преобразований получим

$$E[P^*] = H + P_l - \frac{H - t_{\text{RTS}}}{p_c} Z_N(0) - p_c^{-1} \int_0^{P_l} Z_N(x) dx, \quad (6.8)$$

где

$$Z_N(x) = \{1 - \tau[F(P_l) - F(x)]\}^N - (1 - \tau)^N - N\tau[F(x) + 1 - F(P_l)](1 - \tau)^{N-1}.$$

Тем самым мы определили средние продолжительности всех типов виртуальных слотов.

Аналогично [137] рассмотрим временной интервал t_v между двумя последовательными успешными передачами. Этот интервал, называемый виртуальным временем передачи, представляет собой время между двумя последовательными окончаниями интервалов DIFS. Тогда пропускная способность в условиях насыщения определяется следующим образом:

$$S = \frac{V_c E[P]}{E[t_v]}, \quad (6.9)$$

где $E[t_v]$ — среднее значение t_v . В общем случае виртуальное время передачи t_v может состоять из $l = 1, 2, \dots$ виртуальных слотов, где последний слот «успешный», $k = 0, \dots, l-1$ — «коллизсионные» слоты и $l-1-k$ — «пустые», т.е. $t_v = T_s + kT_c + (l-1-k)\sigma$. Тогда

$$E[t_v] = p_s \sum_{l=1}^{\infty} \sum_{k=0}^{l-1} [T_s + kT_c + (l-1-k)\sigma] \left(\frac{l-1}{k} \right) p_c^k p_e^{l-1-k}.$$

Вычислив эту сумму, получим

$$E[t_v] = T_s + \frac{p_c}{p_s} T_c + \frac{p_e}{p_s} \sigma. \quad (6.10)$$

Такое же выражение было получено в [137].

Как следует из формул (6.3), (6.9) и (6.10), для нахождения пропускной способности S в условиях насыщения остается определить вероятность передачи τ . Чтобы решить данную задачу, модифицируем алгоритм, предложенный в [137], с тем чтобы обобщить значения параметров правила отсрочки и учесть пороговое значение числа повторных попыток R . Приняв во внимание геометрическое распределение времени отсрочки b , имеем $\tau = 1/(E[b] + 1)$, где $E[b]$ — среднее значение времени отсрочки. Так как значения b выбираются из набора $\{0, \dots, w-1\}$, $E[b]$ определяется через среднее значение $E[w]$: $E[b] = (E[w] - 1)/2$. Более подробно последнее равенство доказывается в лемме 2 в [137]. Отсюда,

$$\tau = 2/(E[w] + 1), \quad (6.11)$$

где нужно найти среднее значение конкурентного окна $E[w]$. Станция может находиться в одном из R состояний, где номер состояния i ($i = 0, \dots, R-1$) равен текущему числу попыток передачи. Из состояния i , в котором значение конкурентного окна определяется (6.2), станция переходит в состояние 0, если передача была успешной, а если произошла коллизия, то в состояние $(i+1) \bmod R$. Так как вероятность неудачной передачи равна

$$p = 1 - (1 - \tau)^{n-1}, \quad (6.12)$$

то вероятность отказа определяется формулой

$$p_{rej} = p^R = [1 - (1 - \tau)^{n-1}]^R.$$

Лемма 1. Пусть число попыток передачи текущего пакета n_τ ограничено величиной R . Тогда средняя величина конкурентного окна определяется следующими

формулами: в случае $R \leq m$

$$E[w] = \left\{ \sum_{i=1}^R p^{i-1} W_i + \frac{p^R W_R - W_0}{1-p} \right\} \frac{(1-p)^2}{1-p^R} \quad (6.13)$$

и

$$E[w] = \left(\sum_{i=1}^m p^{i-1} W_i + W_m \sum_{i=m}^{R-1} (i-m+2) p^i + \frac{(R-m+1) W_m p^R - W_0}{1-p} \right) \frac{(1-p)^2}{1-p^R}, \quad (6.14)$$

в обратном случае.

Доказательство. Рассмотрим процесс передачи пакета с момента начала передачи до ее успешного завершения или до момента, когда число попыток достигнет предела R . В этом временном промежутке может произойти $i = 0, 1, \dots, R$ коллизий. Вероятность случая i , когда произошло i коллизий, равна:

$$\mathcal{P}_i^{coll} = (1-p)p^i \text{ при } i = 0, 1, \dots, R-1 \text{ и } \mathcal{P}_R^{coll} = p^R \text{ при } i = R.$$

Пусть $\bar{W}_i$ — сумма конкурентных окон, подсчитанная в течение всего рассматриваемого процесса в случае i . Тогда средний размер конкурентного окна (в расчете на одну попытку передачи) равен

$$E[w] = \left\{ \sum_{i=0}^{R-1} \mathcal{P}_i^{coll} \bar{W}_i + \mathcal{P}_R^{coll} \bar{W}_{R-1} \right\} / \left\{ \sum_{i=0}^{R-1} (i+1) \mathcal{P}_i^{coll} + R \mathcal{P}_R^{coll} \right\}. \quad (6.15)$$

Приняв во внимание (6.2), находим значения $\bar{W}_i$:

$$\bar{W}_i = \sum_{j=0}^i W_j = W_{i+1} - W_0, \quad i < m, \quad (6.16)$$

$$\bar{W}_i = (i-m+1)W_m + \sum_{j=0}^{m-1} W_j = (i-m+2)W_m - W_0, \quad i \geq m. \quad (6.17)$$

Так как знаменатель (6.15) равен $(1-p^R)/(1-p)$, искомые выражения могут быть найдены путем подстановки (6.16) и (6.17) в числитель (6.15).

Решив систему уравнений (6.11)–(6.14), найдем вероятность передачи τ и, следовательно, пропускную способность S в условиях насыщения.

6.8 Эффект захвата

6.8.1 Описание эффекта захвата

В обобщенной модели, рассмотренной в предыдущем разделе, считалось, что любая станция может начать передачу в начале каждого слота. Однако в реальной схеме DCF это не совсем так.

Как видно из рис. 6.18 и 6.19, отсчет времени отсрочки ведется, когда канал простаивает. Станция начинает передачу, когда значение счетчика времени отсрочки достигает нуля. Отсчет слотов отсрочки останавливается, когда канал занят, а возобновляется, только когда канал свободен в течение отрезка времени

$\sigma + \text{DIFS}$ (если предыдущая передача была успешной) или $\sigma + \text{DIFS}$ (при неудачной предыдущей передаче). Рассмотрим слот, следующий сразу за интервалом DIFS, завершающим успешную передачу пакета станции А. В начале этого слота значение счетчика отсрочки станции А равно b , а счетчики остальных станций остаются на тех же значениях, что и до начала передачи станцией А. Таким образом, этот слот является *неконкурентным*: в течение его может вести передачу только станция А, если ее время отсрочки b оказывается равным 0 (в [212] мы назвали эту ситуацию *мгновенным повтором* передачи). Соответственно, попытки передачи, выполняемые в результате мгновенного повтора, назовем *мгновенно повторяемыми попытками*, отличая их от остальных, *обычных*, попыток. Таким образом, станция А может провести целую серию передач, мгновенно повторяя их, причем ни одна из этих мгновенно повторяемых попыток передачи не испытывает коллизии ввиду отсутствия конкуренции со стороны остальных станций. Аналогично в начале слота, следующего непосредственно после интервала EIFS, завершающего коллизию нескольких станций, только эти станции могут передавать, мгновенно повторяя свои попытки. Кроме того, после завершения успешной передачи конкурентное окно данной станции имеет минимальное значение и равно $CW_{\min}$, т.е. рассматриваемая станция имеет больше шансов на выигрыш соревнования за канал. В этом неравноправии доступа к общему каналу и заключается эффект захвата, отмеченный в [134] и исследованный в [212].

Чтобы вычислить пропускную способность в условиях насыщения, используем то же выражение (6.9) из предыдущего раздела. Таким образом, задача состоит в оценке среднего значения $E[t_v]$ виртуального времени передачи t_v между двумя соседними успешными передачами с учетом эффекта захвата. Данный эффект изменяет поведение станции, только что успешно завершившей передачу (для определенности положим, что это станция 1), по сравнению с другими станциями, причем станция 1 имеет больше шансов на захват канала передачи, чем другие станции. Поэтому, чтобы выделить данную станцию из числа остальных, назовем ее *привилегированной* станцией (или станцией, находящейся в *привилегированном состоянии*), а остальные станции — *обычными* станциями (станциями, находящиеся в *обычном состоянии*).

6.8.2 Оценка виртуального времени передачи

Отсчет виртуального времени передачи может быть сразу же завершён при успешной передаче привилегированной станции (ПС) с вероятностью W_0^{-1} . Это происходит, когда выбранное время отсрочки ПС равно 0. В противном случае, после «пустого» слота отсрочки σ ПС включается в соревнование с остальными станциями за предоставление канала. Обозначим эту оставшуюся часть виртуального времени передачи через t_v^c . Тогда

$$E[t_v] = W_0^{-1}T_s + (1 - W_0^{-1})(\sigma + E[t_v^c]), \quad (6.18)$$

где $E[t_v^c]$ — среднее значение рассматриваемого интервала t_v^c . При этом учтем вероятность мгновенного повтора передачи после коллизии и пренебрежем возможностью коллизии при мгновенном повторе, т.е. будем считать, что мгновенно повторяемая попытка всегда успешна. Данное предположение основывается на

следующем: 1) только станции, участвовавшие в исходной коллизии, могут быть вовлечены в коллизиию мгновенно повторяемых попыток (повторную коллизиию); 2) в большинстве случаев конкурентные окна данных станций удваиваются после исходной коллизии. Поэтому вероятность повторной коллизии, по меньшей мере, в несколько раз меньше, чем вероятность возникновения исходной коллизии. Будем различать О-коллизии, в которых участвуют лишь пакеты обычных станций, и П-коллизии, в которых также участвует ПС.

Так же, как и в разделе 6.7, разделим интервал t_v^c на виртуальные слоты, исходя из принципа, что все обычные станции могут попытаться осуществить передачу в начале каждого виртуального слота. В таком случае имеем следующие типы виртуальных слотов.

- «Пустой» слот отсрочки σ , в котором ни одна из станций не пытается осуществить передачу.
- Слот обычной станции (ОС-слот), в котором лишь одна из обычных станций осуществляет передачу или одна из станций, участвовавших в О-коллизии, осуществляет успешную мгновенную повторную попытку передачи.
- Слот привилегированной или обычной станции (ПОС-слот), в котором передачу осуществляет только ПС или после П-коллизии незамедлительно следует успешная мгновенная попытка повторной передачи.
- Слоты О-коллизии (ОК-слот) и П-коллизии (ПК-слот), сразу же за которыми не следует мгновенная попытка повторной передачи, причем после каждого из этих слотов следует «пустой» слот отсрочки σ , завершающий коллизиию.

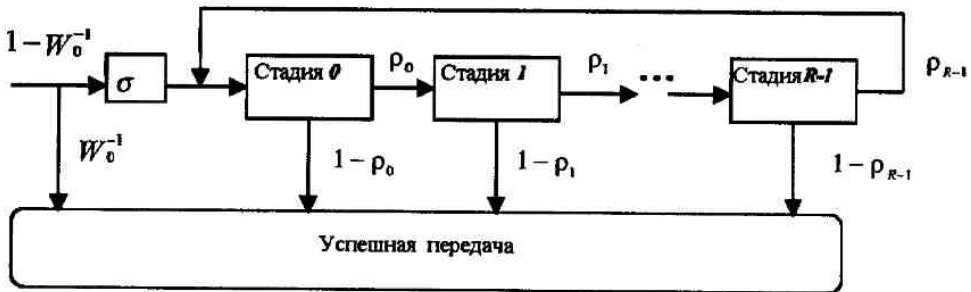


Рис. 6.21. Эволюция состояния сети в течение интервала виртуальной передачи

Рассмотрим процесс изменения состояния БЛС, изображенный на рис. 6.21. БЛС находится в данном слоте на стадии j ($j = 0, 1, \dots, R-1$), если в начале слота ПС находилась на стадии j . Начало «конкурентного» интервала t_v^c означает, что ПС, находившаяся на стадии 0, выбрала время отсрочки $b_0 > 0$. БЛС переходит на стадию 0 и остается на ней в течение виртуальных слотов $k = 1, \dots, b_0$ (см. рис. 6.22), при этом ПС осуществляет передачу только во время слота b_0 при условии, что ни одна из обычных станций не смогла осуществить успешную передачу в течение предыдущих $b_0 - 1$ слотов. Если ни один из b_0 виртуальных

слотов не был успешным, БЛС переходит на стадию 1, и т. д. Если БЛС перешла на стадию $R-1$ и все b_{R-1} слотов оказались неуспешными, она возвращается на стадию 0.

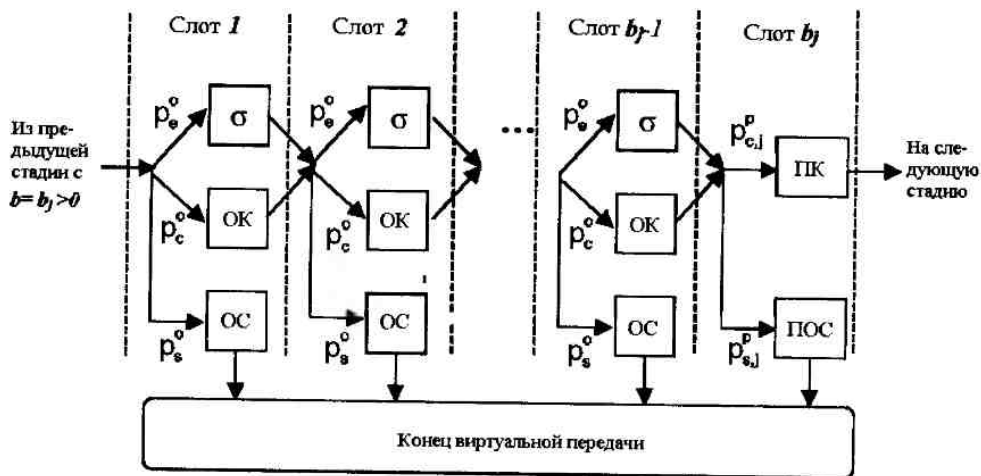


Рис. 6.22. Эволюция на стадии j

При описании поведения обычной станции мы считаем, что: 1) в начале каждого слота обычные станции осуществляют передачу с вероятностью τ ; 2) сразу же после коллизии обычная станция повторяет попытку с вероятностью p_{rep} .

Подробно рассмотрим поведение БЛС на некоторой стадии j (см. рис. 6.22). Так как переход на данную стадию означает, что ПС не предпринимает попытку передачи сразу же после коллизии (когда $b_j = 0$), время отсрочки ПС b_j выбирается с равной вероятностью из набора значений $\{1, \dots, W_j - 1\}$. Во время первых $b_j - 1$ виртуальных слотов передавать могут только рядовые станции, т. е. каждый из этих слотов может быть ОС-слотом, «пустым» слотом или ОК-слотом с вероятностями

$$p_s^o = \hat{p}_s^o + \sum_{k=2}^{N-1} \binom{N-1}{k} \tau^k (1-\tau)^{N-1-k} k p_{rep} (1-p_{rep})^{k-1} = \quad (6.19)$$

$$= (N-1)\tau[p_{rep}(1-\tau p_{rep})^{N-2} + (1-p_{rep})(1-\tau)^{N-2}],$$

$$p_e^o = (1-\tau)^{N-1}, \quad p_c^o = 1 - p_e^o - p_s^o, \quad (6.20)$$

где $\hat{p}_s^o = (N-1)\tau(1-\tau)^{N-2}$ — это вероятность того, что передавать будет лишь одна станция. Пусть ни один из первых $b_j - 1$ виртуальных слотов не был успешным. В этом случае, условная вероятность которого равна

$$\pi_j = (W_j - 1)^{-1} \sum_{i=1}^{W_j-1} (p_e^o + p_c^o)^{i-1} = \frac{1 - (p_e^o + p_c^o)^{W_j-1}}{(W_j - 1)p_s^o},$$

ПС осуществляет передачу в слоте b_j . Поэтому слот b_j может быть ПОС- или ПК-слотом с условными вероятностями $p_{s,j}^p$ и $p_{c,j}^p = 1 - p_{s,j}^p$, где первая определяется

следующим образом:

$$\begin{aligned} p_{s,j}^p &= \hat{p}_s^p + \sum_{k=1}^{N-1} \binom{N-1}{k} \tau^k (1-\tau)^{N-1-k} (1-p_{rep})^k \left[W_{h_j}^{-1} + \frac{(1-W_{h_j}^{-1})k p_{rep}}{1-p_{rep}} \right] = \\ &= (1-W_{h_j}^{-1})(1-\tau)^{N-1} + \left[W_{h_j}^{-1} + (1-W_{h_j}^{-1}) \frac{(N-1)\tau p_{rep}}{1-p_{rep}} \right] (1-\tau p_{rep})^{N-1}, \end{aligned} \quad (6.21)$$

где $\hat{p}_s^p = (1-\tau)^{N-1}$ — вероятность того, что только ПС осуществляет передачу, а $h_j = (j+1)_{mod R}$ — номер стадии отсрочки, соседней с j . Чтобы получить выражения (6.19)-(6.21) (и для дальнейшего анализа), мы пренебрегли вероятностью возникновения повторных коллизий (т.е. коллизий, которые возникают сразу же после первой коллизии), так как отношение вероятности повторных коллизий $\beta_{r/i}$ к вероятности первичных коллизий довольно мало: $\beta_{r/i} \approx p_{rep}^2$ в случае О-коллизий и $\beta_{r/i} \approx W_{h_j}^{-1} p_{rep} + (1-W_{h_j}^{-1})(N-2)p_{rep}^2 \tau$ в случае П-коллизий (эти приближенные оценки верны при $N\tau \ll 1$).

Таким образом, БЛС переходит со стадии j на следующую стадию (см. рис. 6.21) с вероятностью

$$\rho_j = \pi_j p_{c,j}^p. \quad (6.22)$$

Пусть L_j и $\bar{L}_j$ ($j = 0, 1, \dots, R-1$) — средние значения продолжительности соответственно стадии j , завершающейся ОС-слотом или ПОС-слотом, и неуспешной стадии j , за которой следует переход к следующей стадии. Проанализировав схему на рис. 6.21, получим:

$$\begin{aligned} E[t_v^c] &= \sum_{j=0}^{R-1} (1-\rho_j) Q_j \sum_{r=0}^{\infty} (L_j + \bar{Q}_j + r \bar{Q}_R) Q_R^r = \\ &= \sum_{j=0}^{R-1} \frac{(1-\rho_j) Q_j}{1-Q_R} \left\{ L_j + \bar{Q}_j + \frac{Q_R \bar{Q}_R}{1-Q_R} \right\}, \end{aligned} \quad (6.23)$$

где

$$\bar{Q}_j = \sum_{i=0}^{j-1} \bar{L}_i, \quad Q_j = \prod_{i=0}^{j-1} \rho_i, \quad j = 1, \dots, R, \quad \bar{Q}_0 = 0, \quad Q_0 = 1.$$

Аналогичным образом найдем среднее число отказов пакетов ПС за виртуальный интервал передачи:

$$\begin{aligned} N_{rej}^p &= (1-W_0^{-1}) \sum_{r=0}^{\infty} Q_R^r \left\{ r \sum_{j=0}^{R-1} (1-\rho_j) Q_j + (p_{s,R-1}^p - \hat{p}_s^p) \pi_{R-1} Q_{R-1} \right\} = \\ &= \frac{1-W_0^{-1}}{1-Q_R} [Q_R + (p_{s,R-1}^p - \hat{p}_s^p) \pi_{R-1} Q_{R-1}]. \end{aligned} \quad (6.24)$$

Остается найти значения L_j и $\bar{L}_j$ для всех $j = 0, 1, \dots, R-1$. Начнем с определения средних размеров слотов для всех типов виртуальных слотов. На любой стадии длительность «пустого» слота равна σ , а средние значения ОК-, ОС- и

ПК-слотов равны

$$\bar{t}_{OC} = T_c^O + \sigma, \quad \bar{t}_{OS} = T_s + T_c^O \frac{p_s^o - \hat{p}_s^o}{p_s^o} \quad \text{и} \quad \bar{t}_{POC} = T_c^P + \sigma$$

соответственно. Средняя длительность ПОС-слота зависит от номера шага j :

$$\bar{t}_{POS}^{(j)} = T_s + T_c^P \frac{p_{s,j}^p - \hat{p}_s^p}{p_{s,j}^p}.$$

В этих формулах T_c^O и T_c^P — это средние продолжительности О- и П-коллизий. T_c^O и T_c^P вычисляются аналогично T_c в разделе 6.7, т. е. по формулам (6.5) и (6.8) при условии, что (6.8) преобразуется следующим образом: p_c и $Z_N(x)$ замещаются на $1 - p_e^o - \hat{p}_s^o$ и $Z_{N-1}(x)$ при вычислении T_c^O или на $1 - \hat{p}_s^p$ и

$$\hat{Z}_{N-1}(x) = [F(x) + 1 - F(P_l)]\{[1 - \tau(F(P_l) - F(x))]^{N-1} - \hat{p}_s^p\}$$

при вычислении T_c^P .

Используя полученные средние длительности виртуальных слотов, проанализируем рис. 6.22, на котором изображена эволюция состояния БЛС на стадии j , и найдем искомые значения L_j и $\bar{L}_j$:

$$L_j = T_s + \frac{p_s^o}{1 - \rho_j} \left\{ \sum_{i=2}^{W_j-1} g_j(i) + p_{s,j}^p g_j(W_j) + (p_s^o - \hat{p}_s^o)(1 - \pi_j) T_c^O / p_s^o + (p_{s,j}^p - \hat{p}_s^p) \pi_j T_c^P \right\}, \quad (6.25)$$

$$\bar{L}_j = g_j(W_j) / \pi_j + T_c^P + \sigma, \quad (6.26)$$

где

$$g_j(i) = \frac{\bar{t}_{ce}}{W_j - 1} \sum_{h=2}^{i-1} (h-1)(p_e^o + p_c^o)^{h-1}, \quad \bar{t}_{ce} = \sigma + \frac{p_c^o}{p_e^o + p_c^o} T_c^O.$$

Таким образом, подставив сперва полученные значения L_j и $\bar{L}_j$ в (6.23), а затем значение $E[t_v]$ в (6.18), найдем среднюю длительность виртуальной передачи $E[t_v]$ а следовательно (из (6.9)), и пропускную способность S в условиях насыщения. Тем не менее, чтобы применить данные формулы, необходимо знать значения параметров τ и p_{rep} , определяющих поведение обычных станций.

6.8.3 Оценка вероятностей передачи и мгновенного повторения попытки передачи для обычных станций

Поведение станции, находящейся в обычном состоянии, изображено на рис. 6.23. При переходе из привилегированного состояния процесс начинается с j -го шага отсрочки с вероятностью η_j . Станция переходит на следующую стадию отсрочки h_j , если текущий шаг j завершает коллизия, сразу же за которой не следует повторная попытка передачи. Вероятность такой коллизии равна

$$q_j = (1 - W_{h_j}^{-1})[1 - (1 - \bar{q})(1 - \tau)^{N-2}], \quad (6.27)$$

где $\bar{q}$ — это вероятность того, что текущая передача данной обычной станции вступит в коллизия с передачей ПС. Поэтому необходимо оценить вероятности $\bar{q}$ и η_j , $j = 0, \dots, R-1$, перед тем как найти τ и $p_{\text{пер}}$.

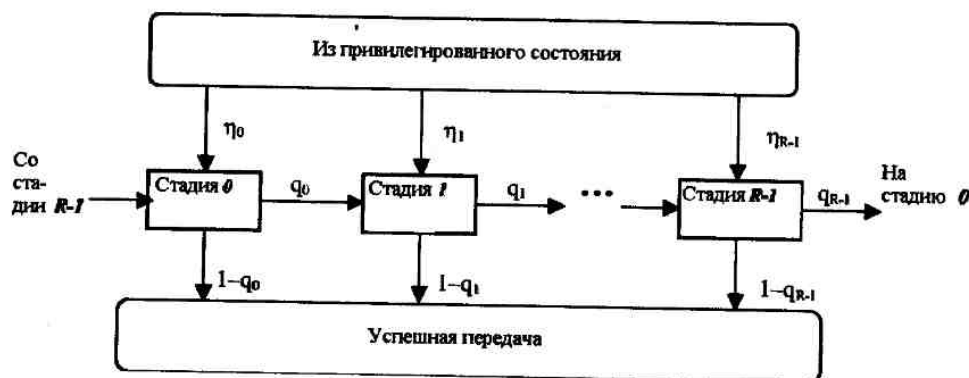


Рис. 6.23. Эволюция обычной станции

В качестве оценки $\bar{q}$ мы можем использовать долю виртуальных слотов, когда ПС ведет передачу, т. е.

$$\bar{q} = N_{\text{slot}}^p / N_{\text{slot}}, \quad (6.28)$$

где N_{slot}^p и N_{slot} — это средние значения количества этих слотов и всех виртуальных слотов соответственно, вычисляемые в течение виртуального времени передачи t_v . Аналогично (6.23) имеем:

$$N_{\text{slot}}^p = \sum_{j=0}^{R-1} \frac{(1-\rho_j)Q_j}{1-Q_R} \left\{ \frac{\pi_j p_{s,j}^p}{1-\rho_j} + j + \frac{RQ_R}{1-Q_R} \right\}, \quad (6.29)$$

$$N_{\text{slot}} = \sum_{j=0}^{R-1} \frac{(1-\rho_j)Q_j}{1-Q_R} \left\{ n_j + \sum_{i=0}^{j-1} \bar{n}_i + \frac{Q_R}{1-Q_R} \sum_{i=0}^{R-1} \bar{n}_i \right\}, \quad (6.30)$$

где n_j и $\bar{n}_j$ — средние количества слотов на успешных и неуспешных стадиях $j = 0, \dots, R-1$, которые находятся аналогично L_j and $\bar{L}_j$:

$$n_j = 1 + \frac{1}{(1-\rho_j)\bar{t}_{ce}} \left\{ p_s^o \sum_{i=3}^{W_j-1} g_j(i) + p_{s,j}^p g_j(W_j) \right\},$$

$$\bar{n}_j = g_j(W_j) / (\pi_j \bar{t}_{ce}) + 1.$$

Таким образом, $\bar{q}$ определена.

Рассмотрим случай, когда данная станция находится в привилегированном состоянии на шаге j перед передачей другой станции, что вызывает переход данной станции в обычное состояние. Этот переход может произойти в следующих двух случаях:

- 1) случай А. Другой станции удается осуществить передачу в данном слоте при отсутствии передачи данной станции;

- 2) случай В. Происходит П-коллизия с мгновенной повторной попыткой другой станции, участвовавшей в коллизии, и без мгновенной повторной передачи данной станции.

Вероятности случаев А и В на успешной стадии j равны $\xi_j^A/(1-\rho_j)$ и $\xi_j^B/(1-\rho_j)$, где $\xi_j^A = 1 - \pi_j$ и (см. (6.21))

$$\xi_j^B = \pi_j(1 - W_{h_j}^{-1})(N-1)\tau p_{rep}(1 - \tau p_{rep})^{N-2}. \quad (6.31)$$

После перехода в обычное состояние рассматриваемая станция остается на том же стадии j в случае А и на следующей стадии h_j в случае В. Следовательно, эволюция обычной станции начинается со стадии отсрочки j с вероятностью

$$\eta_j = \left\{ Q_j \xi_j^A + Q_{\nu_j} \xi_{\nu_j}^B \right\} / \sum_{i=0}^{R-1} Q_i (\xi_i^A + \xi_i^B), \quad (6.32)$$

где ν_j — номер стадии, предшествующей j , т. е. $\nu_j = j-1$ при $j > 0$ и $\nu_0 = R-1$.

Очевидно,

$$\tau = 1/(E[b^*] + 1), \quad (6.33)$$

где $E[b^*]$ — среднее количество виртуальных слотов (в расчете на одну попытку обычной станции, не являющуюся мгновенной повторной попыткой), в которые обычная станция воздерживается от передачи. Значения $E[b^*]$ и вероятность мгновенного повторения передачи p_{rep} , а также вероятность отказа находятся при помощи следующей леммы.

Лемма 2. Пусть станция находится в обычном состоянии. Тогда значения $E[b^*]$, вероятности мгновенной повторной передачи p_{rep} и среднего числа N_{rej}^o повторных отказов данной станции (вычисляемого с момента перехода данной станции в обычное состояние до успешной передачи) получаются при помощи следующих формул:

$$E[b^*] = \sum_{j=0}^{R-1} \eta_j \left\{ \sum_{i=0}^{R-1} v_{ij} \left(a_{ij} + \frac{AV}{1-V} \right) \right\} / \left\{ \sum_{i=0}^{R-1} v_{ij} \left(i + 1 + \frac{RV}{1-V} \right) \right\}, \quad (6.34)$$

$$p_{rep} = \sum_{j=0}^{R-1} \eta_j \left\{ \sum_{i=0}^{R-1} v_{ij} \chi_{(j+i) \bmod R} \right\} / \left\{ \sum_{i=0}^{R-1} v_{ij} \left(i + \chi_{(j+i) \bmod R} + \frac{RV}{1-V} \right) \right\}, \quad (6.35)$$

$$N_{rej}^o = \sum_{j=0}^{R-1} \sum_{i=0}^{R-1} \frac{\eta_j v_{ij}}{1-V} \left(\frac{V}{1-V} + \gamma_{ij} \right), \quad (6.36)$$

где

$$A = \sum_{h=0}^{R-1} \frac{W_h - 2}{2}, \quad V = \prod_{h=0}^{R-1} q_h, \quad v_{0j} = 1 - q_j, \quad a_{0j} = \frac{\xi_j^A Q_j b_{0j}^* + \frac{W_j-2}{2} Q_{\nu_j} \xi_{\nu_j}^B}{Q_j \xi_j^A + Q_{\nu_j} \xi_{\nu_j}^B};$$

$$b_{0j}^* = \left[\left(W_j - 1 - \frac{2}{p_s^o} \right) \frac{W_j - 2}{2} + \frac{1}{p_s^o} \left(\frac{1 - (1 - p_s^o)^{W_j-1}}{p_s^o} - 1 \right) \right] / [(W_j - 1)(1 - \pi_j)], \quad (6.37)$$

$$\begin{aligned}
 v_{ij} &= (1 - q_i) \prod_{h=j}^{i+j-1} q_h, \quad a_{ij} = a_{0,j} + \sum_{h=j+1}^{i+j} \frac{W_h - 2}{2} \quad \text{при } i < R - j; \\
 v_{ij} &= (1 - q_i) \left(\prod_{h=j}^{R-1} q_h \right) \left(\prod_{h=0}^{i+j-R-1} q_h \right), \\
 a_{ij} &= a_{0,j} + \sum_{h=j}^{R-1} \frac{W_h - 2}{2} + \sum_{h=0}^{i+j-R} \frac{W_h - 2}{2} \quad \text{при } R - j \leq i < R;
 \end{aligned}$$

вероятность q_j была определена в (6.27) для всех $j = 0, \dots, R-1$; $\chi_j = q_j / [(W_{h_j} - 1) \times (1 - q_j)]$ — условная вероятность того, что данная станция успешно осуществит передачу на шаге j ; и наконец, $\gamma_{ij} = 1$ при $i \geq R - j$, $\gamma_{ij} = \chi_{R-1}$ при $i + j = R - 1$ и $\gamma_{ij} = 0$ в остальных случаях.

Доказательство. Рассмотрим эволюцию станции, изображенную на рис. 6.23, с момента, когда данная станция переходит в обычное состояние, до момента, когда данная станция успешно осуществит передачу. В процессе данной эволюции будем подсчитывать число B^* виртуальных слотов, когда данная обычная станция воздерживается от передачи, число коллизий N_c рассматриваемой станции и число отказов передачи ее пакетов N_r . Вероятность случая $\mathcal{C}(j, i, r)$ равна

$$\hat{P}[\mathcal{C}(j, i, r)] = \eta_j v_{ij} V^r.$$

В этом случае последней стадией с ненулевым временем отсрочки является $(j+i) \bmod R$. Сразу же после перехода в обычное состояние на стадии j согласно случаю А станция продолжает откладывать передачу до следующей попытки в течение

$$b_{0j}^* = \frac{p_s^c}{(W_j - 1)(1 - \pi_j)} \sum_{i=2}^{W_j-1} \sum_{k=2}^i (i-k)(1 - p_s^c)^{k-2}$$

виртуальных слотов (вычислив сумму, легко получим (6.37)). Если станция находится на стадии j в результате коллизии или перехода в обычное состояние согласно случаю В, тогда она откладывает передачу в течение в среднем $b_j^* = (W_j - 2)/2$ виртуальных слотов, потому что, во-первых, время отсрочки выбирается с равной вероятностью из $\{1, \dots, W_j - 1\}$ без мгновенного повтора и, во-вторых, первый слот отсрочки не засчитывается (он входит в тот же виртуальный слот, что и последняя попытка).

Поэтому в случае $\mathcal{C}(j, i, r)$ имеем $B^* = a_{ij} + rA$, $N_c = \chi_{(j+i) \bmod R} + i + rR$ и $N_r = r + \gamma_{ij}$. Следовательно,

$$E[b^*] = \sum_{j=0}^{R-1} \sum_{i=0}^{R-1} \sum_{r=0}^{\infty} (a_{ij} + rA) \hat{P}[\mathcal{C}(j, i, r)] / \sum_{j=0}^{R-1} \sum_{i=0}^{R-1} \sum_{r=0}^{\infty} (i + 1 + rR) \hat{P}[\mathcal{C}(j, i, r)]$$

— вероятность того, что за коллизией данной обычной станции следует незамедлительная передача, — равна

$$p_{rep} = \frac{\sum_{j=0}^{R-1} \sum_{i=0}^{R-1} \sum_{r=0}^{\infty} \chi_{(j+i) \bmod R} \hat{P}[\mathcal{C}(j, i, r)]}{\sum_{j=0}^{R-1} \sum_{i=0}^{R-1} \sum_{r=0}^{\infty} (i + \chi_{(j+i) \bmod R} + rR) \hat{P}[\mathcal{C}(j, i, r)]},$$

и среднее количество отказов передачи пакета данной станции определяется формулой

$$N_{rej}^o = \sum_{j=0}^{R-1} \sum_{i=0}^{R-1} \sum_{r=0}^{\infty} (r + \gamma_{ij}) \hat{P}[\mathcal{C}(j, i, r)].$$

Вычисляя суммы для r , получим искомые формулы (6.34)–(6.36).

Чтобы найти вероятность отказов пакетов, используем следующее выражение:

$$p_{rej} = N_{rej} / (N_{rej} + 1), \quad (6.38)$$

где N_{rej} — среднее число отказов между последовательными моментами успешной передачи. Очевидно,

$$N_{rej} = N_{rej}^p + (1 - W_0^{-1}) \left[\sum_{i=0}^{R-1} Q_i(\xi_i^A + \xi_i^B) \right] N_{rej}^o, \quad (6.39)$$

где N_{rej}^p и N_{rej}^o были определены при помощи (6.24) и леммы 2.

Зная $\tilde{q}$ и η_j , $j = 0, \dots, R-1$ и решив систему уравнений (6.27) и (6.34)–(6.35), найдем значения τ и p_{rep} . Таким образом, чтобы оценить пропускную способность S в условиях насыщения, используем следующий итеративный метод.

- Шаг 0. Задать начальные значения $\tau = \tau^0$ и $p_{rep} = p_{rep}^0$.
- Шаг 1. Вычислить значения $\tilde{q}$ и η_j , $j = 0, \dots, R-1$ при помощи формул (6.28)–(6.32).
- Шаг 2. Найти измененные значения τ^1 and p_{rep}^1 при помощи формул (6.33)–(6.35).
- Шаг 3. Сравнить τ^1 с τ^0 и p_{rep}^1 с p_{rep}^0 . Если разность значений превышает некоторый заранее определенный предел, вернуться к шагу 1, выбрав $(\tau^0 + \tau^1)/2$ и $(p_{rep}^0 + p_{rep}^1)/2$ в качестве начальных значений.
- Шаг 4. Вычислить сначала вероятность отказов по формулам (6.24), (6.38) и (6.39), затем средние длительности шагов L_j и $\bar{L}_j$ по формулам (6.25) и (6.26), затем среднее время виртуальной передачи $E[t_v]$ по формулам (6.18) и (6.23) и, в конце концов, пропускную способность S в условиях насыщения при помощи уравнения (6.9).

Мы не приводим доказательство сходимости данного метода из-за его достаточно большой сложности. Интуитивно понятно, что уравнение (6.32) имеет единственное решение, так как увеличение вероятности передачи τ ведет к увеличению вероятности возникновения коллизии, следовательно, к увеличению среднего числа слотов $E[b^*]$, в которых обычные станции предпринимают попытки.

На практике многочисленные примеры применения данного метода с различными значениями параметров БЛС показывают, что данный метод обеспечивает быструю сходимость к искомому решению и высокую скорость вычисления значения показателей производительности. Чтобы вычислить S и p_{rej} при помощи данного метода, вычислительной машине с процессором Intel Celeron 400 MHz требуется менее одной секунды.

Будем далее называть аналитические модели, разработанные в разделах 6.7 и 6.8, начальной моделью (Н-моделью) и моделью эффекта захвата (ЭЗ-модель) соответственно.

6.9 Численные результаты оценки производительности в условиях идеального канала

Для подтверждения работоспособности Н-модели и ЭЗ-модели полученные с их помощью результаты сравнивались с результатами имитационного моделирования, полученными на основе системы GPSS (General Purpose Simulation System) [195], разработанной Гордоном из IBM. На данный момент существует несколько различных реализаций GPSS с различным синтаксисом и семантикой (см., например, GPSS World [157]).

Нами рассматривалась БЛС под управлением протокола IEEE 802.11 с механизмом DCF с применением технологии FHSS на физическом уровне. Значения параметров протокола, использованные при получении численных результатов в случае как аналитического, так и имитационного моделирования, приняты по умолчанию в [222] для физического уровня FHSS и приводятся в табл. 6.2.

Таблица 6.2. Значения параметров протокола, использованные при получении численных результатов

Скорость канала (V_c)	1 Мбит/с	Слот (σ)	50 мкс
MAC+PHY-заголовок ($V_c H$)	400 бит	SIFS	28 мкс
ACK ($V_c t_{ACK}$)	240 бит	DIFS	128 мкс
Задержка распространения сигнала (δ)	1 мкс	EIFS	796 мкс

На рис. 6.24 представлены некоторые результаты исследования для следующего случая:

- число станций $N = 20$,
- применяется только механизм базового доступа,
- максимальное конкурентное окно $CW_{\max} = 1024$,
- размер пакета данных (в байтах) выбирается с равной вероятностью из $\{0, 1, \dots, 1250\}$.

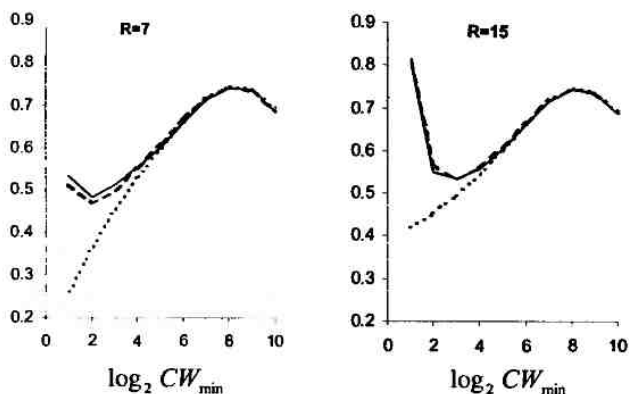


Рис. 6.24. Пропускная способность S (Мбит/с) в зависимости от минимального конкурентного окна

Результаты представлены на рис. 6.24 в виде графиков $S(CW_{\min})$, отображающих зависимость пропускной способности в условиях насыщения от минимального размера конкурентного окна $CW_{\min}$ в логарифмическом масштабе. Сплошная, пунктирная и точечная линии соответствуют результатам, полученным при помощи имитационного моделирования, модели эффекта захвата и начальной модели соответственно.

Сравнение графиков на рис. 6.24 показывает, что обе модели довольно точно соответствуют действительности для большого диапазона значений $CW_{\min}$: ошибки начальной модели и модели эффекта захвата не превышают 4 и 1% соответственно при $CW_{\min} \geq 16$. При уменьшении значения $CW_{\min}$ ниже 16 эффект захвата вызывает увеличение ошибки начальной модели и делает ее непригодной при $CW_{\min} = 2$, в то время как ошибка модели эффекта захвата даже в этом предельном случае остается небольшой и не превышает 7%.

Таблица 6.3. Вероятность отказа p_{rej} в зависимости от минимального конкурентного окна $CW_{\min}$

$CW_{\min}$	2	4	8	16	32	64
ЭЗ-модель	0,2424	0,2195	0,1052	0,0387	0,0142	0,0040
Н-модель	0,7012	0,3302	0,1240	0,0413	0,0145	0,0040
Имит. модел.	0,3082	0,2145	0,1019	0,0387	0,0142	0,0041

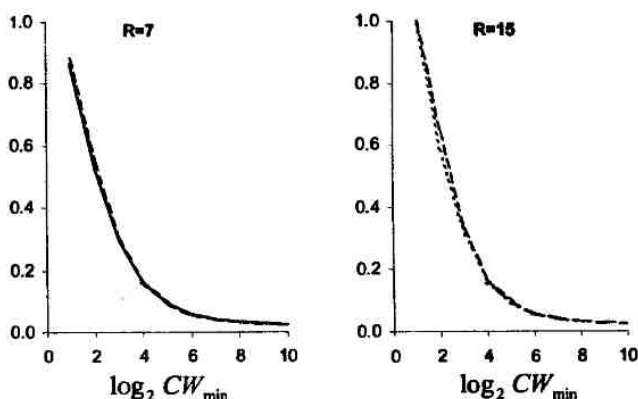
Некоторые результаты оценки вероятности отказа p_{rej} с $n = 50$ и $R = 7$ приведены в табл. 6.3. Как видим, начальная модель работает хорошо только при $CW_{\min} \geq 16$ (когда ошибка не превышает 3%) и становится неэффективной при меньших значениях $CW_{\min}$. В противоположность начальной модели ошибка метода эффекта захвата не превышает 3% для всех значений $CW_{\min}$, кроме $CW_{\min} = 2$, когда ошибка достигает 20% вследствие довольно высокой вероятности возникновения повторных коллизий. В частности, при $CW_{\min} = 16$, рекомендуемым в [222] для БЛС с FHSS PPHU, ошибки для начальной модели и модели эффекта захвата равны 6,9 и 0,1% соответственно.

Вернемся к рис. 6.24. Как показывают модель эффекта захвата и результаты имитационного моделирования, каждая кривая $S(CW_{\min})$ имеет два локальных

максимума при $CW_{\min} \hat{=} 2$ и $CW_{\min} = CW_{\min}^{opt}$ ($CW_{\min}^{opt} = 256$ на рис. 6.24). Характер кривых не очень сильно изменяется с изменением количества станций N и введением механизма RTS/CTS для передачи всех пакетов или только их части. Увеличение числа станций N приводит к увеличению $CW_{\min}^{opt}$, а применение RTS/CTS механизма делает график более плавным.

В случае, когда предел R , ограничивающий количество попыток, предпринимаемых для передачи пакета, равен значению $R = 7$, принятому по умолчанию в [222], максимальное значение пропускной способности в условиях насыщения достигается при $CW_{\min}^{opt} = 256$, т.е. оптимальное значение $CW_{\min}$ существенно больше, чем рекомендуемое значение $CW_{\min} = 16$. Тот же вывод получен в исследованиях, проведенных в [130, 137]. Поэтому в [134, 137] были предложены некоторые механизмы в качестве надстройки над DCF в целях фактического увеличения стандартного значения $CW_{\min}$.

Рис. 6.25. Вероятность захвата в зависимости от минимального конкурентного окна



При $R = 15$ пропускная способность в условиях насыщения максимальна при $CW_{\min} = 2$: $S(2) = 0,801$, в то время как $S(256) = 0,744$. Более того, если увеличить $CW_{\max}$ совместно с R , значение $S(2)$ станет очень близким к значению максимальной пропускной способности в условиях насыщения, равной $E[P]V_c/T_s$. Этот факт, являющийся следствием эффекта захвата, становится наиболее очевидным при $CW_{\min} = 2$: одна из станций захватывает канал на длительное время, в то время как остальные станции находятся на стадиях отсрочки с большими значениями конкурентного окна, поэтому практически не участвуют в соревновании за канал. Таким образом, если принимать во внимание лишь значение пропускной способности, тогда в условиях насыщения следующая стратегия доступа является оптимальной: $CW_{\min} = 2$ и максимальные значения R и $CW_{\max}$. Однако захват канала на долгое время одной станцией не справедлив по отношению к другим станциям.

В качестве показателя этой несправедливости используем вероятность захвата P_{seize} , определяемую следующим образом. Пусть станция А успешно завершила передачу пакета. Тогда следующей станцией, которая успешно передаст пакет, будет станция А с вероятностью захвата P_{seize} . Рассматривая эволюцию БЛС на

рис. 6.21 и 6.22, получаем аналогично (6.29):

$$P_{seize} = W_0^{-1} + (1 - W_0^{-1}) \sum_{j=0}^{R-1} \frac{\pi_j Q_j}{1 - Q_R} [(1 - W_{h_j}^{-1})(1 - \tau)^{N-1} + W_{h_j}^{-1}(1 - \tau p_{rep})^{N-1}]. \quad (6.40)$$

Кривые $P_{seize}(CW_{min})$, изображенные на рис. 6.25, соответствуют кривым $S(CW_{min})$ на рис. 6.24. Как и на рис. 6.24, результаты, полученные при помощи имитационного моделирования и рассчитанные по формуле (6.40), показаны соответственно сплошными и штриховыми кривыми. Сравнивая эти кривые, видим, что приближение (6.40) является очень точным при любых значениях CW_{min} : сплошная и пунктирная кривые практически совпадают. При $R = 15$ и $CW_{min} = 2$ показатель несправедливости распределения канала почти равен единице: $P_{seize} = 0,991$, в то время как при $CW_{min} = CW_{min}^{opt}$ несправедливость мала: $P_{seize} = 0,033$. Таким образом, теперь мы можем определить оптимальные параметры протокола, с тем чтобы максимизировать пропускную способность при ограничении несправедливости, т.е. ее индекс не должен превышать заданный предел $\bar{P}_{seize}$: $P_{seize} \leq \bar{P}_{seize}$. Численные результаты показывают, что CW_{min}^{opt} является оптимальным значением CW_{min} при $\bar{P}_{seize} < 0,98$.

6.10 Производительность локальных сетей при наличии помех

В предыдущих разделах производительность БЛС оценивалась в предположении идеального канала, т.е. в отсутствие помех. Однако пренебрежение помехами обычно приводит к существенному завышению оценок пропускной способности, так как в современных городских условиях электромагнитные помехи — неизбежный фактор, ухудшающий пропускную способность сети из-за искажения пакетов. Поэтому этот и следующий разделы посвящены влиянию помех на производительность сетей и оценке эффективности таких вспомогательных механизмов протокола IEEE 802.11, как RTS/CTS и фрагментация пакетов. Кроме того, в данном разделе предлагается и исследуется модификация правила отсрочки, которая обещает значительно повысить пропускную способность сети, по крайней мере в случае, когда помехи представляют собой так называемый «белый шум».

При стандартном правиле отсрочки (см. раздел 6.6) конкурентное окно удваивается не только после коллизии, но и после передачи кадра, искаженного помехами, что, очевидно, может привести к непроизводительным потерям времени (особенно в случае редких коллизий). К сожалению, станция не может распознать причину неудачной попытки передачи: коллизия или помехи. Однако можно заметить, что вероятность искажения помехами коротких кадров RTS, CTS и ACK намного меньше, чем длинных кадров DATA. Поэтому, используя механизм RTS/CTS, станция с достаточной степенью уверенности может считать, что причиной неудачной передачи кадра RTS является коллизия, в то время как передача кадра DATA, предвояемая последовательностью RTS-CTS (так же, как и передача кадра DATA, не являющегося первым в цепочке фрагментов и по-

этому следующего за кадром ACK), происходит из-за искажения его помехами. Назовем такой кадр DATA *резервируемым* кадром, отражая тот факт, что время, требуемое для его передачи, резервируется предшествующим кадром CTS или ACK.

Поэтому в дополнение к стандартному правилу отсрочки будем изучать также следующую его модификацию: конкурентное окно w устанавливается равным W_0 и счетчик n_r обнуляется после неудачной передачи резервируемого кадра DATA (так как передававшая его станция считается победителем в соревновании за канал).

Как и в [130, 137], при анализе схемы DCF мы применим следующее допущение: все станции БЛС изменяют значения своих счетчиков отсрочки непосредственно после интервала DIFS или EIFS, завершающего попытку передачи, т. е. передававшая станция (или станции в случае коллизии) изменяет свое конкурентное окно и выбирает случайным образом начальное значение счетчика отсрочки из множества $(0, \dots, w-1)$, в то время как остальные станции просто уменьшают свои счетчики отсрочки на единицу. (В реальности, как показано в предыдущем разделе, остальные станции смогут сделать это только спустя слот σ с момента окончания DIFS или EIFS.) Таким образом, в начале каждого слота любая станция может начать свою передачу. Согласно численным результатам, приведенным в конце предыдущего раздела, это допущение не влияет значительно на результаты оценки пропускной способности при значениях W_0 , рекомендованных в [222] (т. е. при $W_0 = 16$ для технологии FHSS или $W_0 = 32$ для технологии DSSS).

6.10.1 Оценка пропускной способности

При анализе БЛС в данном и следующем разделах мы считаем, что все станции используют одни и те же значения RTS-порога и порога фрагментации ($\bar{P}$ и L_f), а вероятностное распределение $\{d_\ell, \ell = 1, \dots, \ell_{\max}\}$ длин пакетов, выбираемых станцией из очереди, одинаково для всех станций. Так как расстояния между станциями БЛС считаются малыми, мы также предполагаем, что помехи проявляются одновременно на всех станциях, т. е. все станции «слышат» общий беспроводный канал одинаково.

Аналогично разделу 6.7 все время работы исследуемой БЛС разобьем на неоднородные виртуальные слоты, каждый из которых с вероятностями p_e , p_s и p_c , определяемыми (6.3), может быть «пустым» (p_e), «успешным» (p_s) или «коллизийным».

В общем случае каждая попытка передачи пакета представляет собой попытку передать цепочку кадров DATA (содержащих последовательные фрагменты этого пакета и перемежаемых ответными кадрами ACK, а также короткими межкадровыми промежутками SIFS), причем первый кадр DATA из любой цепочки содержит первый фрагмент из числа тех, которые еще не были успешно переданы. В частности, на рис. 6.20 показаны две такие цепочки, причем передача второй цепочки обусловлена искажением кадра DATA со вторым фрагментом. Как было уже описано в разделе 6.6, передаваемые цепочки разделены интер-

валами отсрочки. Передача кадра DATA с фрагментом, являющимся первым в цепочке и превышающим по длине предел $\bar{P}$, предваряется посылкой кадра RTS и получением разрешающего кадра CTS.

Свяжем каждую попытку передачи с парой (ℓ, k) , где ℓ — длина пакета, к которой относится эта цепочка, а $k + 1$ — число фрагментов этого пакета, которые осталось передать, причем $k = 0, K(\ell)$, где $K(\ell)$ — целая часть отношения $(\ell - 1)/L_f$. Обозначим также через $r_{\ell k}^1$ и r_{ℓ}^0 длины первого и последнего фрагментов остатка пакета, передаваемого при данной попытке. Для примера рассмотрим попытку передачи пакета длиной $\ell = 4, 5L_f$, считая, что первые два фрагмента этого пакета уже успешно переданы, а передача третьего фрагмента была неудачной при предыдущей попытке. Тогда цепочка, передаваемая при данной попытке, включает в себя кадры DATA с двумя фрагментами длиной L_f и последним фрагментом длиной $L_f/2$, и эти фрагменты образуют остаток пакета. Кроме того, рассматриваемая попытка будет связана с парой $(\ell = 4, 5L_f, k = 2)$, причем $r_{\ell k}^1 = L_f$ и $r_{\ell}^0 = L_f/2$.

Пусть вероятность того, что произвольно выбранная попытка передачи связана с конкретной парой (ℓ, k) , равна $\hat{d}_{\ell k}$.

Искомая пропускная способность S определяется как среднее число бит полезной информации, переданных за секунду. Заметим, что мы должны подсчитывать эти биты только после успешного завершения целого пакета, но не отдельного его фрагмента, так как процесс передачи пакета может завершиться отказом несмотря на то, что некоторые из его фрагментов могли быть успешно переданы. Поэтому S находим по формуле:

$$S = p_s \sum_{\ell=\ell_{\min}}^{\ell_{\max}} \sum_{k=0}^{K(\ell)} 8\ell\pi_h(\ell, k)\hat{d}_{\ell k}/[p_e\sigma + p_sT_s + p_cT_c], \quad (6.41)$$

где T_s и T_c — средние длительности «успешного» и «коллизийного» слотов (т.е. знаменатель представляет собой среднюю длительность виртуального слота), а $\pi_h(\ell, k)$ — вероятность того, что попытка, связанная с парой (ℓ, k) , выполнена успешно и, таким образом, завершила передачу пакета длиной ℓ .

Длительность «коллизийного» слота складывается из времени передачи кадра максимальной длины из числа кадров, вовлеченных в коллизии, плюс интервал EIFS, который всегда завершает неудачную попытку передачи. Тогда при j кадрах, вовлеченных в коллизии, ее длительность равна либо $t_{\text{RTS}} + \text{EIFS} + \delta$ с вероятностью D_0^j , где

$$D_0 = \sum_{h=\bar{P}+1}^{L_f} d_r^*,$$

либо $t_d(r) + \text{EIFS} + \delta$, $r = 1, \dots, \min(L_f, \bar{P})$, с вероятностью $D_r^j - D_{r-1}^j$, где

$$D_r = D_0 + \sum_{h=1}^r d_r^*,$$

$t_d(r) = H + 8r/V$ — время передачи кадра DATA, включающего фрагмент длиной r байт; V — скорость канала (в бит/с). Кроме того, d_r^* — вероятность того, что при данной попытке передается цепочка, длина первого фрагмента которой

равна τ , т. е.

$$d_r^* = \varphi(r) = \sum_{k=0}^{k_{\max}(r)} \hat{d}_{kL_f+r, 0}, \quad r < L_f, \quad d_{L_f}^* = 1 - \sum_{\ell=1}^{\ell_{\max}} \hat{d}_{\ell, 0} + \varphi(L_f),$$

где $k_{\max}(r)$ — целая часть отношения $(\ell_{\max} - r)/L_f$. Кроме того, здесь и далее используется булева функция $\mathbf{1}[\text{cond}]$, принимающая единичное значение при выполнении условия cond .

Так как вероятность того, что в коллизии вовлечено ровно j кадров, равна

$$\pi_j = p_c^{-1} \binom{N}{j} \tau^j (1 - \tau)^{N-j},$$

то средняя длительность «коллизийного» слота определяется выражением

$$T_c = \text{EIFS} + \delta + \sum_{j=2}^N \pi_j \left\{ t_{\text{RTS}} D_0^j + \sum_{r=1}^{\min(L_f, \bar{P})} t_d(r) (D_r^j - D_r^j) \right\}$$

или после преобразований

$$\begin{aligned} T_c = & \text{EIFS} + \delta + p_c^{-1} t_{\text{RTS}} [(1 - \tau + \tau D_0)^N - p_e - D_0 p_s] + \\ & + p_c^{-1} \sum_{r=1}^{\min(L_f, \bar{P})} t_d(r) [(1 - \tau + \tau D_r)^N - (1 - \tau + \tau D_{r-1})^N - d_r^* p_s]. \end{aligned} \quad (6.42)$$

Теперь рассмотрим «успешный» слот. В начале «успешного» слота одна и только одна станция предпринимает попытку передачи, которая с вероятностью $\hat{d}_{\ell k}$ связана с парой (ℓ, k) . Эта попытка завершится успехом для данной пары (ℓ, k) (т. е. успешной передачей всего пакета длиной ℓ), если ни один из кадров, передаваемых в ходе этой попытки, не искажен помехами, т. е. с вероятностью

$$\pi_h(\ell, k) = Q(r_{\ell k}^1) [1 - \xi(L_f)]^k [1 - \xi(r_{\ell}^0)], \quad (6.43)$$

где

$$Q(\ell_f) = 1 - \xi_{rc} = (1 - \xi_r)(1 - \xi_a)$$

при $\ell_f > \bar{P}$ и $Q(\ell_f) = 1$ при $\ell_f \leq \bar{P}$,

$$\xi(\ell_f) = 1 - [1 - \xi_d(\ell_f)](1 - \xi_a),$$

наконец, $\xi_d(\ell_f)$, ξ_r , и ξ_a — вероятности искажения помехами кадра DATA с фрагментом длиной ℓ_f ($\xi_d(\ell)$), кадра RTS (ξ_r) и кадров CTS и ACK (ξ_a) одинакового формата [222]. (То есть ξ_{rc} и $\xi(\ell_f)$ — вероятности искажения помехами соответственно последовательностей RTS-CTS и DATA-ACK.) Эти вероятности искажения определяются на основе показателя BER (Bit Error Rate) — вероятности искажения одного бита (этот показатель будем называть также интенсивностью помех), т. е. кадр, состоящий из f байт, искажается с вероятностью $\xi_f = 1 - \exp\{-f\text{BER}\}$. (В этом разделе рассматривается только случай постоянного значения BER, по крайней мере, в течение процесса передачи одного пакета.)

Средняя длительность попытки передачи остатка пакета, описываемого парой (ℓ, k) , очевидно, равна

$$\begin{aligned}
 t_s(\ell, k) = & Q(r_{\ell k}^1) \sum_{i=0}^k (t_{d,i} + \text{SIFS} + \delta) [1 - \xi(L_f)]^i + \\
 & + Q(r_{\ell k}^1) (t_{\text{ACK}} + \text{SIFS} + \delta) \sum_{i=0}^k (1 - \xi_{d,i}) [1 - \xi(L_f)]^i + \\
 & + 1(r_{\ell k}^1 > \bar{P}) [t_{\text{RTS}} + \delta + \text{SIFS} + (1 - \xi_r)(t_{\text{CTS}} + \delta + \text{SIFS})] - \\
 & - \text{SIFS} + \pi_h(\ell, k) \text{DIFS} + [1 - \pi_h(\ell, k)] \text{EIFS},
 \end{aligned} \tag{6.44}$$

где

$$t_{d,i} = \begin{cases} t_d(r_{\ell}^0), & i = k \\ t_d(L_f), & i < k, \end{cases} \quad \xi_{d,i} = \begin{cases} \xi_d(r_{\ell}^0), & i = k \\ \xi_d(L_f), & i < k. \end{cases}$$

Следовательно, средняя длительность T_s «успешного» слота равна

$$T_s = \sum_{\ell=1}^{\ell_{\max}} \sum_{k=0}^{K(\ell)} t_s(\ell, k) \hat{d}_{\ell k}. \tag{6.45}$$

Таким образом, искомая пропускная способность S находится по формулам (6.41)–(6.45) при условии, что известны вероятность начала передачи τ и вероятностное распределение $\{\hat{d}_{\ell k}\}$.

6.10.2 Вероятности начала передачи и отказа

Рассмотрим процесс передачи пакета длиной ℓ некоторой станцией, начинающийся в момент выбора пакета из очереди и заканчивающийся либо его успешной передачей, либо отказом. Пусть f_{ℓ} и $\bar{w}_{\ell}$ — средние количества соответственно попыток, совершаемых этой станцией, и виртуальных слотов, в которые станция воздерживается от передачи, в течение рассматриваемого процесса (назовем $\bar{w}_{\ell}$ суммарным временем отсрочки). Тогда

$$\tau = \sum_{\ell=1}^{\ell_{\max}} d_{\ell} f_{\ell} / \sum_{\ell=1}^{\ell_{\max}} d_{\ell} (f_{\ell} + \bar{w}_{\ell}). \tag{6.46}$$

Кроме того, найдем также усредненную вероятность отказа $\bar{p}_{rej}$, происходящего при достижении одним из счетчиков n_s или n_d своего предела (N_s или N_d). Эта вероятность определяется следующей суммой:

$$\bar{p}_{rej} = \sum_{\ell=1}^{\ell_{\max}} d_{\ell} p_{rej}(\ell), \tag{6.47}$$

где $p_{rej}(\ell)$ — вероятность отказа для пакета длиной ℓ .

Передача коротких нефрагментированных пакетов. Начнем с более простого случая $\ell \leq \max\{\bar{P}, L_f\}$, когда пакет не делится на фрагменты и передается механизмом базового доступа. (Заметим, что в этом случае стандартное и модифицированное правила отсрочки совпадают.) При этом число попыток i

ограничено N_s , поэтому

$$f_\ell = \sum_{i=1}^{N_s} i\psi_\ell(i), \quad \bar{w}_\ell = \sum_{i=1}^{N_s} \bar{W}_i\psi_\ell(i), \quad (6.48)$$

где $\psi_\ell(i)$ — вероятность, что происходит ровно i попыток. Суммарное время отсрочки $\bar{W}_i$, определенное при условии i попыток, находится с помощью (6.2): при $1 \leq i \leq m+1$

$$\bar{W}_i = \sum_{k=0}^{i-1} \frac{W_k - 1}{2} = W_{i-1} - \frac{W_0 + i}{2}, \quad (6.49)$$

а при $i > m+1$

$$\begin{aligned} \bar{W}_i &= \sum_{k=0}^m \frac{W_k - 1}{2} + \frac{W_m - 1}{2}(i - 1 - m) = \\ &= W_m \frac{i - m + 1}{2} - \frac{W_0 + i}{2}. \end{aligned} \quad (6.50)$$

Найдем вероятности $\psi_\ell(i)$, $i = 1, \dots, N_s$. Вероятность неудачной попытки равна

$$\pi_{cd}(\ell) = 1 - (1 - \pi_c)[1 - \xi(\ell)],$$

где $\pi_c = 1 - (1 - \tau)^{N-1}$ — вероятность коллизии текущей попытки. Тогда процесс передачи пакета завершится успешно на i -й попытке с вероятностью

$$\psi_\ell^s(i) = [1 - \pi_{cd}(\ell)][\pi_{cd}(\ell)]^{i-1}, \quad i = 1, \dots, N_s \quad (6.51)$$

или закончится отказом после N_s -й попытки с вероятностью

$$p_{rej}(\ell) = [\pi_{cd}(\ell)]^{N_s}. \quad (6.52)$$

Следовательно,

$$\psi_\ell(i) = \psi_\ell^s(i) = [1 - \pi_{cd}(\ell)][\pi_{cd}(\ell)]^{i-1}, \quad i = 1, \dots, N_s - 1, \quad (6.53)$$

$$\psi_\ell(N_s) = \psi_\ell^s(N_s) + p_{rej}(\ell) = [\pi_{cd}(\ell)]^{N_s-1}. \quad (6.54)$$

Передача длинных нефрагментированных пакетов. Теперь пусть $L_f \geq \ell > \bar{P}$. В этом случае число передач кадра DATA ограничено N_d , и каждой такой передаче может предшествовать $0, \dots, N_s - 1$ неудачных передач кадра RTS.

Пусть в процессе передачи пакета длиной ℓ происходит ровно i_d и i_r повторных передач соответственно кадров DATA и RTS. Так как вероятности неудачных передач кадров DATA и RTS равны $\xi(\ell)$ и $\pi_{cr} = 1 - (1 - \pi_c)(1 - \xi_{rc})$, то этот процесс завершится успешно (случай S_{i_d, i_r}) с вероятностью

$$\zeta_\ell^s(i_d, i_r) = A_\ell(i_d, i_r)g(i_r, i_d + 1), \quad i_d < N_d, \quad i_r \leq (N_s - 1)(i_d + 1), \quad (6.55)$$

где

$$A_\ell(i_d, i_r) = (1 - \pi_{cr})^{i_d+1} [1 - \xi(\ell)][\xi(\ell)]^{i_d} \pi_{cr}^{i_r},$$

или неудачно (отказом) из-за достижения либо предела N_s (случай $\mathcal{R}_{i_d, i_r}$) с вероятностью

$$\zeta_\ell^r(i_d, i_r) = \frac{A_\ell(i_d, i_r)g(i_r - N_s, i_d)}{(1 - \pi_{cr})[1 - \xi(\ell)]}, \quad i_d < N_d, \quad i_r - N_s \leq (N_s - 1)i_d, \quad (6.56)$$

либо предела N_d (случай $\mathcal{D}_{i_r}$) с вероятностью

$$\zeta_\ell^d(i_r) = A_\ell(N_d - 1, i_r) \frac{\xi(\ell)}{1 - \xi(\ell)} g(i_r, N_d), \quad i_r \leq (N_s - 1)N_d. \quad (6.57)$$

В (6.55)–(6.57) целочисленная функция $g(u, v)$ означает число способов, которыми u неразличимых шаров (неудачных передач RTS) могут быть помещены в v урн (интервалов, предшествующих каждой из передач DATA) так, чтобы каждая урна содержала не более $N_s - 1$ шаров. Эта функция вычисляется рекурсивно:

$$g(0, v) = 1 \quad \forall v \geq 0, \quad g(u, 1) = \begin{cases} 1, & u < N_s \\ 0, & u \geq N_s, \end{cases}$$

$$g(u, v) = \sum_{k=0}^{\min(u, N_s-1)} g(u-k, v-1), \quad v \geq 2, \quad u > 0.$$

Таким образом, вероятность осуществления ровно i попыток равна

$$\psi_\ell(i) = \psi_\ell^s(i) + \mathbf{1}(i \geq N_s) \sum_{i_d=0}^{\min\{i-N_s, N_d-1\}} \zeta_\ell^r(i_d, i - i_d) + \mathbf{1}(i \geq N_d) \zeta_\ell^d(i - N_d),$$

$$i = 1, \dots, i_m^1, \quad (6.58)$$

где

$$\psi_\ell^s(i) = \sum_{i_d=0}^{\min\{i, N_d\}-1} \zeta_\ell^s(i_d, i - i_d - 1) \quad (6.59)$$

— вероятность успешного завершения процесса на i -й попытке.

Теперь находим среднее число попыток f_ℓ с помощью первого уравнения (6.48), в котором N_s заменяется на $i_m^1 = N_s N_d$. Кроме того, вероятность отказа для пакета длиной ℓ равна

$$p_{rej}(\ell) = \sum_{i_d=0}^{N_d-1} \sum_{i_r=N_s}^{N_s+i_d(N_s-1)} \zeta_\ell^r(i_d, i_r) + \sum_{i_r=0}^{N_d(N_s-1)} \zeta_\ell^d(i_r). \quad (6.60)$$

Чтобы завершить анализ этого случая, осталось найти суммарное время отсрочки $\bar{w}_\ell$ для всех ℓ , и для этого придется учитывать отличие стандартного и модифицированного правил отсрочки. При *стандартном правиле отсрочки* счетчик n_r возрастает на единицу при каждой неудаче, поэтому $\bar{w}_\ell$ определяется вторым уравнением (6.48), в котором N_s заменяется на i_m^1 .

Рассмотрим *модифицированное правило отсрочки*, при котором после неудачной передачи резервируемого кадра DATA конкурентное окно w сбрасывается к W_0 , а счетчик n_r обнуляется. При этом правиле имеем:

$$\begin{aligned} \bar{w}_\ell = & \sum_{i_d=0}^{N_d-1} \sum_{i_r=0}^{(i_d+1)(N_s-1)} W_{i_r, i_d}^s \zeta_\ell^s(i_d, i_r) + \\ & + \sum_{i_d=0}^{N_d-1} \sum_{i_r=N_s}^{N_s+i_d(N_s-1)} W_{i_r, i_d}^r \zeta_\ell^r(i_d, i_r) + \sum_{i_r=0}^{N_d(N_s-1)} W_{i_r}^d \zeta_\ell^d(i_r), \end{aligned} \quad (6.61)$$

где W_{i_r, i_d}^s , W_{i_r, i_d}^r и $W_{i_r}^d$ — суммарные времена отсрочки, вычисленные для случаев соответственно $\mathcal{S}_{i_d, i_r}$, $\mathcal{R}_{i_d, i_r}$ и $\mathcal{D}_{i_r}$.

Анализируя случай $\mathcal{S}_{i_d, i_r}$, получаем:

$$W_{i_r, i_d}^s = \sum_{j=1}^{i_d+1} \sum_{i=0}^{\min\{i_r, N_s-1\}} \varphi(j, i, i_d, i_r) \frac{W_i - 1}{2}, \quad i_d < N_d, \quad i_r \leq (N_s - 1)(i_d + 1), \quad (6.62)$$

где $\varphi(j, i, i_d, i_r)$ — вероятность того, что ровно i неудачных передач кадра RTS происходят в промежутке между $(j-1)$ -й и j -й передачами кадра DATA (или до первой передачи кадра DATA, если $j=1$) при условии, что имеет место случай $\mathcal{S}_{i_d, i_r}$. Очевидно,

$$\varphi(j, i, i_d, i_r) = g(i, 1)g(i_r - i, i_d)/g(i_r, i_d + 1),$$

и уравнение (6.62) принимает вид:

$$W_{i_r, i_d}^s = \eta(i_r, i_d) = \frac{i_d + 1}{g(i_r, i_d + 1)} \sum_{i=0}^{\min\{i_r, N_s-1\}} g(i_r - i, i_d) \frac{W_i - 1}{2}. \quad (6.63)$$

Аналогично получаем

$$W_{i_r, i_d}^r = \bar{W}_{N_s+1}(i_d > 0)\eta(i_r - N_s, i_d - 1), \quad i_d < N_d, \quad i_r - N_s \leq (N_s - 1)i_d, \quad (6.64)$$

$$W_{i_r}^d = \eta(i_r, N_d - 1), \quad i_r \leq (N_s - 1)N_d. \quad (6.65)$$

Передача фрагментированных пакетов. В процессе передачи фрагментированного пакета длиной $\ell > L_f$ попытки передачи остатка пакета, описываемые парой (ℓ, k) , $k = 0, \dots, K(\ell) - 1$, могут происходить, только если

1) все предшествовавшие фрагменты $1, \dots, K(\ell) - k$ были успешно переданы, т.е. счетчики n_s и n_ℓ не достигли своих пределов;

2) при первой попытке передачи фрагмента $K(\ell) - k + 1$, являющегося началом остатка, либо кадр DATA, содержащий этот фрагмент, либо ответный кадр АСК был искажен помехами, что привело к увеличению на единицу счетчика n_s или n_d (в зависимости от длины фрагмента).

Вероятность $z_{\ell, k}$ выполнения обоих указанных условий определяется следующими уравнениями: $z_{\ell, K(\ell)} = 1$ и при $k < K(\ell)$

$$z_{\ell, k} = \xi(r_{\ell, k}^1)[1 - p_{rej}(L_f)][1 - \xi(L_f)\hat{p}_{rej}(L_f)]^{K(\ell)-k-1},$$

где $\hat{p}_{rej}(r)$, $r = 1, \dots, L_f$ — вероятность отказа, находящаяся с помощью либо (6.52) с подстановкой $N_s - 1$ вместо N_s при $r \leq \bar{P}$, либо (6.60) с подстановкой $N_d - 1$ вместо N_d в (6.57) и (6.60) при $r > \bar{P}$. (Путем такой подстановки мы учитываем условие 2.)

Кроме того, при $\ell > L_f$ вероятность отказа равна

$$p_{rej}(\ell) = p_{rej}(L_f) + \sum_{k=1}^{K(\ell)-1} \hat{p}_{rej}(L_f) z_{\ell,k} + \hat{p}_{rej}(r_\ell^0) z_{\ell,0}. \quad (6.66)$$

Теперь найдем вероятность $\psi_\ell(i)$ (при $\ell > L_f$), что ровно i попыток происходит в процессе передачи пакета длиной ℓ . Имеем:

$$\psi_\ell(i) = 1[i \leq i_m(L_f)][\psi_{L_f}(i) - \psi_{L_f}^s(i)] + \sum_{h=1}^{\min[i, i_m(L_f)]} \psi_{L_f}^s(h) G_{K(\ell)-1}(\ell, i-h), \quad (6.67)$$

где первое слагаемое (6.67) — это вероятность отказа при передаче первого фрагмента пакета,

$$i_m(r) = \begin{cases} N_s & \text{при } r \leq \bar{P} \\ (N_s - 1)N_d + 1 & \text{при } r > \bar{P}, \end{cases}$$

а $G_k(\ell, j)$ — вероятность того, что происходит ровно j попыток при передаче последних $k+1$ фрагментов при условии, что первые $K(\ell) - k$ фрагментов уже переданы успешно.

Очевидно, $G_k(\ell, j)$ можно вычислить рекурсивно: при $k > 0$

$$\begin{aligned} G_k(\ell, i) = & [1 - \xi(L_f)] G_{k-1}(\ell, i) + 1(0 < i \leq i_m^*(L_f)) \xi(L_f) [\hat{\psi}_{L_f}(i) - \hat{\psi}_{L_f}^s(i)] + \\ & + 1(i > 0) \xi(L_f) \sum_{h=1}^{\min[i, i_{m1}(L_f)]} \hat{\psi}_{L_f}^s(h) G_{k-1}(\ell, i-h) \end{aligned} \quad (6.68)$$

и

$$G_0(\ell, i) = 1(i=0)[1 - \xi(r_\ell^0)] + 1(0 < i \leq i_m^*(r_\ell^0)) \xi(r_\ell^0) \hat{\psi}_{r_\ell^0}(i),$$

где

$$i_m^*(r) = \begin{cases} N_s - 1 & \text{при } r \leq \bar{P} \\ (N_s - 1)(N_d - 1) + 1 & \text{при } r > \bar{P}, \end{cases}$$

а вероятности $\hat{\psi}_r(i)$ и $\hat{\psi}_r^s(i)$ находятся аналогично $\psi_r(i)$ и $\psi_r^s(i)$, т. е. с помощью либо (6.51), (6.53) и (6.54) с подстановкой $N_s - 1$ вместо N_s при $r \leq \bar{P}$, либо (6.58) и (6.59) с подстановкой $N_d - 1$ вместо N_d в (6.57)–(6.59) при $r > \bar{P}$. В (6.68) первое слагаемое определяет условную вероятность того, что при передаче фрагмента $K(\ell) - k + 1$ не происходит попыток (т. е. этот фрагмент не искажается помехами), поэтому все i попыток происходят при передаче последних k фрагментов. Второе слагаемое — это условная вероятность того, что все i попыток происходят при передаче фрагмента $K(\ell) - k + 1$, что приводит к отказу. Последнее слагаемое соответствует случаю, когда фрагмент $K(\ell) - k + 1$ передается успешно после $h > 0$ попыток.

Теперь по формулам (6.48), в которых N_s заменяется на

$$I_m(\ell) = i_m(r_{\ell, K(\ell)}^1) + [(K(\ell) - 1)i_m^*(L_f) + i_m^*(r_\ell^0)],$$

находим среднее число попыток f_ℓ для обоих рассматриваемых правил отсрочки, а также суммарное время отсрочки $\bar{w}_\ell$ для стандартного правила.

При модифицированном правиле отсрочки конкурентное окно w и счетчик n_r сбрасываются к своим начальным значениям после неудачной передачи резервируемого кадра DATA. Это позволяет рассчитывать суммарное время отсрочки отдельно для каждого фрагмента. Следовательно,

$$\bar{w}_\ell = \bar{w}_{L_f} + \sum_{k=1}^{K(\ell)-1} \bar{w}_{L_f}^* z_{\ell k} + \bar{w}_{r_0}^* z_{\ell 0}, \quad (6.69)$$

где $\bar{w}_r^* = \sum_{i=1}^{N_s-1} \bar{W}_i \hat{\psi}_r(i)$, если $r \leq \bar{P}$, а иначе $\bar{w}_r^*$ определяется формулой (6.61), в которой (а также в (6.57) и (6.65)) N_d заменяется на $N_d - 1$.

Таким образом, мы получили алгебраически полную систему уравнений для нахождения значения вероятности начала передачи как при стандартном, так и при модифицированном правиле отсрочки.

Распределение числа попыток. Для использования формул (6.41), (6.42) и (6.45), определяющих значения S , T_c и T_s , осталось найти распределение $\{\hat{d}_{\ell k}\}$. Очевидно,

$$\hat{d}_{\ell k} = d_\ell f_{\ell k} / \sum_{u=1}^{\ell_{\max}} \sum_{v=0}^{K(u)} d_u f_{uv}, \quad \ell = 1, \dots, \ell_{\max}, \quad k = 0, \dots, K(\ell), \quad (6.70)$$

где $f_{\ell k}$ — это среднее число попыток передачи остатка пакета, описываемого парой (ℓ, k) . Для нефрагментированного пакета, когда $\ell \leq L_f$, имеем $f_{\ell 0} = f_\ell$, где f_ℓ определяется первым уравнением (6.48) (с заменой N_s на i_m^1 при $\ell > \bar{P}$), а при $\ell > L_f$ среднее число попыток определяется следующим образом:

$$f_{\ell, K(\ell)} = \sum_{i=1}^{i_m(L_f)} i \psi_{L_f}(i), \quad f_{\ell k} = z_{\ell k} \sum_{i=1}^{i_m^*(r_{\ell k}^1)} i \hat{\psi}_{r_{\ell k}^1}(i), \quad k < K(\ell). \quad (6.71)$$

Итеративная процедура. Оценка вероятности начала передачи τ проводится с помощью следующей итеративной процедуры.

Шаг 0. Определить начальное значение τ .

Шаг 1. Для длин пакетов $\ell = 1, \dots, L_f$ вычислить: вероятности $\psi_\ell(i)$ и $\psi_\ell^s(i)$ для $i = 1, \dots, i_m(\ell)$; вероятность отказа $p_{rej}(\ell)$; среднее число попыток f_ℓ ; суммарное время отсрочки $\bar{w}_\ell$, используя (6.48) и (6.51)–(6.54), если $\ell \leq \bar{P}$. Если $\ell > \bar{P}$, использовать (6.55)–(6.60) и модифицированные уравнения (6.48) с заменой N_s на i_m^1 . (В случае модифицированного правила отсрочки применить (6.61) и (6.63)–(6.65) для оценки $\bar{w}_\ell$.)

Шаг 2. Для всех $\ell = 1, \dots, L_f$ вычислить модифицированные вероятности $\hat{\psi}_\ell(i)$, $\hat{\psi}_\ell^s(i)$ ($i = 1, \dots, i_m^*(\ell)$) и $\hat{p}_{rej}(\ell)$ по тем же формулам, что и для $\psi_\ell(i)$, $\psi_\ell^s(i)$ и $p_{rej}(\ell)$, в которых N_s заменяется на $N_s - 1$ при $\ell \leq \bar{P}$ или N_d заменяется на $N_d - 1$ при $\ell > \bar{P}$.

Шаг 3. Для всех $L_f < \ell \leq \ell_{\max}$ вычислить вероятность $\psi_\ell(i)$, $i = 1, \dots, i_m(\ell)$ по формуле (6.67), среднее число попыток f_ℓ и суммарное время отсрочки $\bar{w}_\ell$ по модифицированным уравнениям (6.48) с заменой N_s на $i_m(\ell)$. (При модифицированном правиле отсрочки вычислить $\bar{w}_\ell$ по формуле (6.69).)

Шаг 4. Используя (6.46), найти модифицированное значение τ и сравнить его с начальным значением. Если их абсолютная разность оказывается большей зара-

нее определенного предела, вернуться на шаг 1, используя полусумму начального и модифицированного значений в качестве нового начального значения.

После этой итеративной процедуры находим усредненную вероятность отказа $\bar{p}_{rej}$ по формулам (6.47), (6.52), (6.60) и (6.66). Наконец, определяем распределение $\{\hat{d}_{ek}\}$ по формулам (6.70)–(6.71) и пропускную способность S по формулам (6.41)–(6.45).

Как и в предыдущем разделе, мы не приводим строгого доказательства сходимости предложенной процедуры.

6.11 Результаты оценки производительности сети с помехами

Объектом численного исследования с применением как аналитического метода, разработанного в предыдущем разделе, так и имитационного моделирования на языке GPSS [195] являлась БЛС, состоящая из N статистически однородных станций, очереди которых насыщены, и работающая под управлением высокоскоростной версии протокола IEEE 802.11 — IEEE 802.11b [161]. Значения параметров протокола, использованные при этом исследовании, приведены в табл. 6.4. Кроме того, размер пакета данных (в байтах) выбирался с равной вероятностью из $\{1, \dots, 1999\}$.

Таблица 6.4. Параметры протокола IEEE 802.11b (короткая преамбула)

Слот, σ	20 мкс
Время распространения сигнала, δ	1 мкс
MAC+PHY-заголовок	49 байт
Время передачи заголовка, H	121 мкс
Длина ACK и CTS	29 байт
Время передачи ACK, t_{ACK}	106 мкс
Длина RTS	35 байт
Время передачи RTS, t_{RTS}	111 мкс
SIFS	10 мкс
DIFS	50 мкс
EIFS	212 мкс
Скорость канала, V	11 Мбит/с
Предел N_s	7
Предел N_d	4
Минимальное конкурентное окно, W_0	32
Максимальное конкурентное окно, W_m	1024

6.11.1 Оптимизация порога RTS и порога фрагментации

На рис. 6.26 и 6.27 представлены некоторые результаты исследования пропускной способности и усредненной вероятности отказа для а) базового доступа и б) ме-

ханизма RTS/CTS (когда соответственно $\bar{P} > l_{\max}$ и $\bar{P} = 0$) при изменении числа станций N . (В этом случае мы не исследуем фрагментацию пакетов и полагаем $L_f = l_{\max}$.) На этих рисунках (как и на рис. 6.29 и 6.30) пунктирные кривые получены с помощью имитационного моделирования, а остальные — с помощью аналитического метода. Прежде всего отметим высокую точность аналитической модели: погрешность никогда не превышает 2% при оценке пропускной способности и 5% при оценке вероятности отказа.

Далее, как можно было ожидать, пропускная способность выше при базовом доступе, если число станций N мало ($N < 30$ на рис. 6.26 и 6.27), в то время как механизм RTS/CTS эффективнее при больших N и обеспечивает примерно одинаковую пропускную способность при изменении N .

Рис. 6.26. Пропускная способность S (Мбит/с) в зависимости от числа станций N при $\text{BER} = 5 \cdot 10^{-5}$ для: а — механизма базового доступа; б — механизма RTS/CTS и с — оптимального гибридного механизма

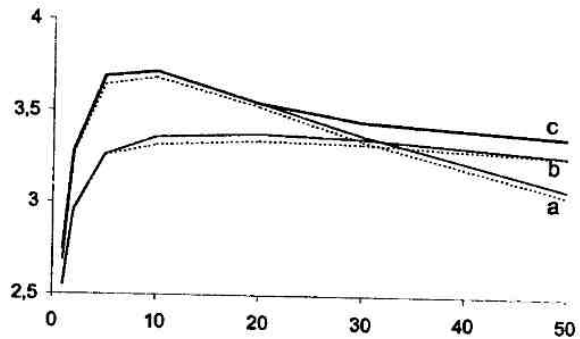
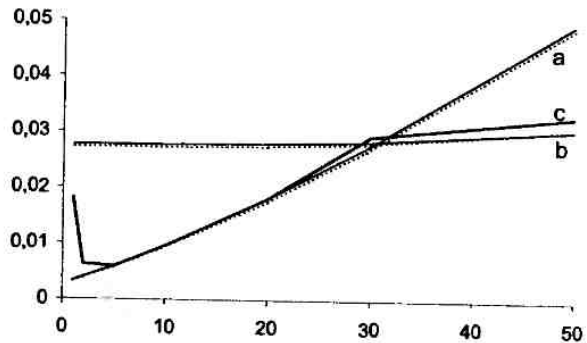


Рис. 6.27. Вероятность отказа $\bar{p}_{rej}$ в зависимости от числа станций N при $\text{BER} = 5 \cdot 10^{-5}$



Яркие кривые на рис. 6.26 и 6.27 получены при гибридном механизме с оптимальным порогом RTS $\bar{P}_{opt}$, обеспечивающим максимальную пропускную способность и зависящим от N . Оптимизирующие кривые показаны на рис. 6.28 для разных BER и были получены аналитически. (Высокая скорость вычислений, обеспечиваемая разработанным методом, позволила находить оптимальные пороги путем полного перебора.) При низких BER (кривая (i)) оптимум $\bar{P}_{opt}$ довольно низок при больших N , возрастает монотонно с уменьшением N до некоторого предела N_b , где $\bar{P}_{opt}$ становится равным $l_{\max} + 1 = 2000$ байтам, и остается постоянным при $N \leq N_b$, т.е. механизм базового доступа является наилучшим при малых N . При больших BER (кривые (ii)–(iv)) график $\bar{P}_{opt}(N)$ немонотонен,

т. е. появляется дополнительный порог $N_b^0 < N_b$ и $\bar{P}_{opt}$ снижается с уменьшением N от N_b^0 до 1. Например, $N_b = 30$ и $N_b^0 = 15$ при $BER = 1 \cdot 10^{-4}$. Оба порога возрастают с ростом BER, но N_b^0 растет быстрее, так что интервал, где механизм базового доступа является наилучшим, исчезает и эти два порога сливаются при очень высоких BER (см. график (iv)).

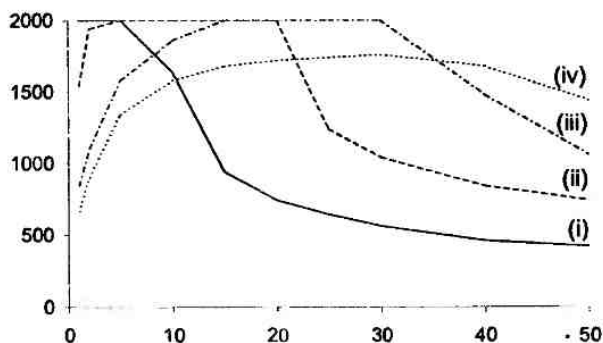


Рис. 6.28. Оптимальный порог RTS $\bar{P}_{opt}$ в зависимости от числа станций N при (i) $BER = 1 \cdot 10^{-5}$, (ii) $BER = 5 \cdot 10^{-5}$, (iii) $BER = 1 \cdot 10^{-4}$, и (iv) $BER = 1.4 \cdot 10^{-4}$

Таблица 6.5. Сравнение базового доступа (БД) и оптимального гибридного механизма (ОГ)

Параметры		S , Мбит/с		$\bar{p}_{rej}$	
BER	N	БД	ОГ	БД	ОГ
$5 \cdot 10^{-5}$	1	2,683	2,742	0,0033	0,0181
$1 \cdot 10^{-4}$	1	1,021	1,296	0,0556	0,1415
$1 \cdot 10^{-4}$	5	1,892	1,936	0,0621	0,1002
$1,4 \cdot 10^{-4}$	1	0,5370	0,7653	0,1474	0,2638
$1,4 \cdot 10^{-4}$	5	1,1333	1,2189	0,1545	0,2163

Таким образом, был получен следующий, весьма неочевидный факт, иллюстрируемый табл. 6.5 (где сравниваются показатели производительности при базовом доступе и оптимальном гибридном механизме): при небольшом числе станций и высоком BER лучшим механизмом является не базовый доступ, а некоторый гибридный механизм, причем повышение пропускной способности, достигаемое путем такой оптимизации, оказывается значительным. Как показывают табл. 6.5 и рис. 6.27, это повышение пропускной способности достигается за счет ухудшения надежности, т. е. повышения вероятности отказа. В рассматриваемом случае вероятность коллизии мала и вероятность неудачи примерно равна вероятности искажения кадра DATA помехами. Поэтому можно считать, что максимальное число попыток передачи пакета равно $N_d = 4$ при передаче пакета механизмом RTS/CTS и равно $N_s = 7$ при Базовом Доступе. Для данных значений длины пакета и BER чем меньше максимальное число попыток, тем больше вероятность отказа, тем меньше среднее время отсрочки и, следовательно, тем выше пропускная способность.

Теперь исследуем влияние BER на пропускную способность S и усредненную вероятность отказа $\bar{p}_{rej}$ в случае применения фрагментации. На рис. 6.29

и 6.30 показаны некоторые результаты этого исследования для случаев передачи: а) нефрагментированных пакетов (графики n/f) и б) фрагментированных пакетов с порогом фрагментации $L_f = 400$ байт (графики f). Яркая кривая на рис. 6.29 показывает результаты для оптимального порога $L_f = L_f^{opt}$, максимизирующего пропускную способность и зависящего от BER. (Здесь использовался порог RTS $\bar{P} = 600$ байт.)

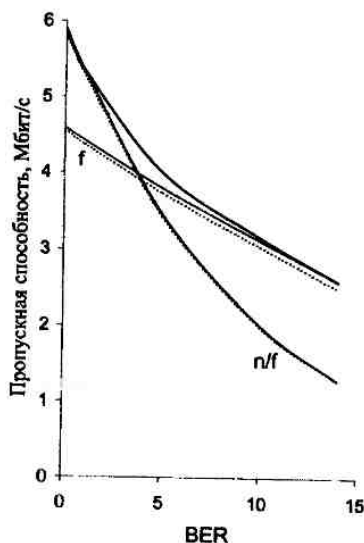


Рис. 6.29. Пропускная способность в зависимости от $BER \cdot 10^5$ при $N = 50$ станциях в случае нефрагментированной (n/f) и фрагментированной (f) передачи

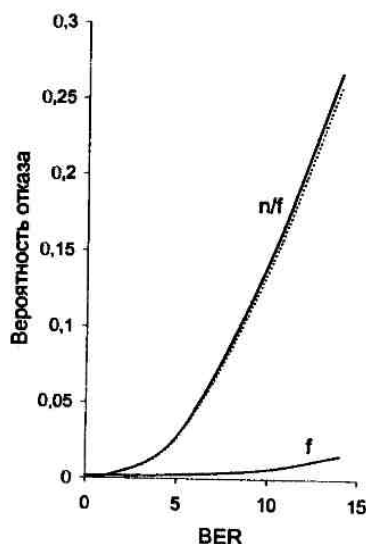


Рис. 6.30. Вероятность отказа в зависимости от $BER \cdot 10^5$

Аналитически определенные оптимальные значения L_f^{opt} изображены графиками на рис. 6.31. Ступенчатый вид этих кривых довольно сложен. С увеличением числа станций область, где нефрагментированная передача оптимальна, расширяется с увеличением числа «ступенек» и их сужением, что является результатом роста вероятности коллизии.

Точка пересечения графиков, относящихся к фрагментированной и нефрагментированной передаче пакетов на рис. 6.29, примерно соответствует порогу $BER = 3,9 \times 10^{-5}$. Для BER, меньших этого порога, фрагментация нежелательна: проигрыш в пропускной способности по сравнению с оптимальной фрагментацией не больше 9%. Для BER, больших этого порога, фрагментация с $L_f = 400$ байт почти оптимальна. Таким образом, хотя (согласно рис. 6.31) оптимальное значение L_f существенно и сложным образом зависит от BER, на практике применима простая стратегия: не фрагментировать, если $BER < \bar{BER}$, и фрагментировать на весьма малые части ($L_f = 400$ байт), если $BER \geq \bar{BER}$.

6.11.2 Сравнительный анализ правил отсрочки

Проведенная оптимизация не вполне корректна, так как мы не учитывали последствия отказа в передаче пакета. В реальности, пакеты, отвергнутые на MAC-

уровне, в большинстве случаев передаются повторно протоколами транспортного уровня (такими, как TCP). Поэтому можно считать, что пакет длины ℓ , который должен быть передан некоторой станцией, попадает в ее MAC-очередь в среднем $1/[1 - p_{rej}(\ell)]$ раз. Введем начальное вероятностное распределение $\{d_\ell^0\}$ для длин пакетов, выбираемых из очереди станции в первый раз. Тогда $\{d_\ell = d_\ell^0/[1 - p_{rej}(\ell)]\}$, и необходимо включить в итеративную процедуру, описанную в конце предыдущего раздела, дополнительный шаг 3а для вычисления вероятностей отказа $p_{rej}(\ell)$ и распределения $\{d_\ell\}$. В следующих далее численных примерах предполагается, что размер пакета, выбираемого из очереди станции в первый раз, равновероятно разыгрывается из множества $\{1, \dots, 1999\}$.

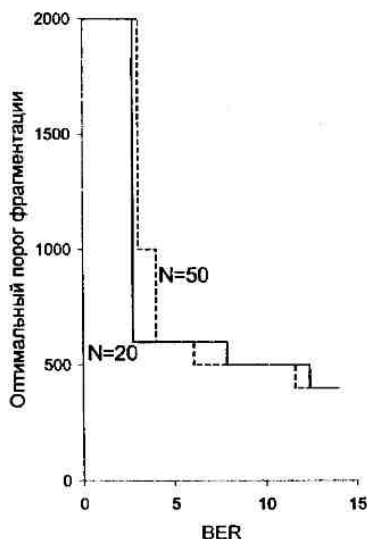


Рис. 6.31. Оптимальный порог фрагментации в зависимости от $BER \cdot 10^5$

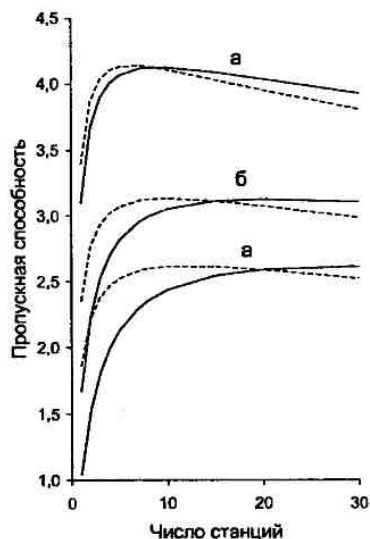


Рис. 6.32. Сравнение правил отсрочки при: а — $BER = 5 \cdot 10^{-5}$, б — $BER = 1 \cdot 10^{-4}$ и в — $BER = 1.4 \cdot 10^{-4}$

Теперь применим разработанный метод для сравнения стандартного и модифицированного правил отсрочки. Результаты этого сравнения показаны на рис. 6.32, где сплошные и штриховые графики соответствуют стандартному и модифицированному правилам и получены с применением оптимальных порогов RTS и фрагментации, которые определялись отдельно для каждого правила с учетом вторичных потоков пакетов, отвергнутых ранее на MAC-уровне.

Видно, что модифицированное правило отсрочки значительно (от 10 до 50%) повышает пропускную способность в случае, когда число станций не слишком велико, а интенсивность помех высока. Стандартное правило отсрочки предпочтительней (относительная разность значений пропускной способности лежит в пределах 5%) при большом числе станций и низком BER. Кроме того, существует и третья область (малое число станций и низкий BER), в которой фрагментация пакетов не используется и механизм базового доступа применяется для всех пакетов. Очевидно, в этой области стандартное и модифицированное правила неразличимы. Эти три области значений BER и N показаны на рис. 6.33 и обо-

значены соответственно через « $M > S$ », « $M < S$ » и « $M = S$ ». Видно, что линия, разделяющая области « $M > S$ » и « $M < S$ », почти прямая, и это дает нам основание заключить, что модифицированное правило отсрочки предпочтительней при $BER > (0,78N - 1,4) \cdot 10^{-5}$.

6.12 Производительность сети при коррелированных сбоях

В предыдущем разделе мы развили методы [130, 137, 212] для исследования влияния помех на производительность локальной сети типа ad hoc, работающей под управлением протокола IEEE 802.11, предполагая сбой канала (т.е. искажения передаваемых пакетов помехами) независимыми, что имеет место, например, в случае канала с аддитивным гауссовым шумом. Однако известно (см., например, [216] и [236]), что поведение беспроводной среды лучше описывается моделью Джилберта [159], представляющей собой цепь Маркова из двух состояний: «хорошего» и «плохого». Эти состояния отличаются значениями BER, предполагаемыми постоянными в каждом из состояний. Очевидно, согласно этой модели сбой канала, вызванные влиянием помех, оказываются коррелированы, и эта корреляция существенно затрудняет аналитическую оценку производительности таких сетей, вынуждая предыдущих исследователей данной проблемы прибегать к имитационному моделированию (см., например, [136]). Тем не менее, как будет показано в данном разделе, производительность таких беспроводных сетей может быть оценена аналитически, если предположить, что времена пребывания канала в «хорошем» и «плохом» состояниях распределены экспоненциально.

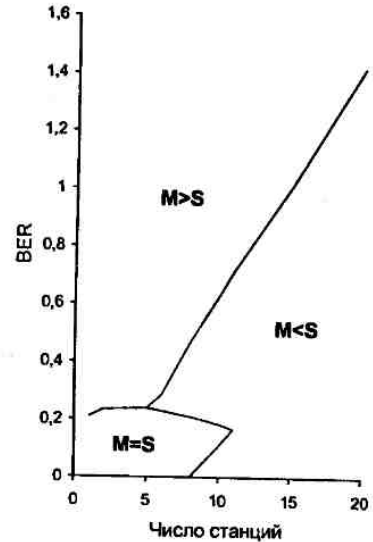


Рис. 6.33. Области эффективности ($BER \cdot 10^{-5}$ в зависимости от числа станций) стандартного и модифицированного правил отсрочки

6.12.1 Пропускная способность при коррелированных сбоях

Итак, рассмотрим ту же БЛС, что и в разделе 6.10, и при тех же допущениях (за исключением коррелированности сбоев). Так как согласно результатам раздела 6.11 (см. также [119, 209]) механизм RTS/CTS эффективен только в редких случаях, когда число активных станций велико (больше 20), в данном разделе предполагается использование метода базового доступа для всех пакетов. Кроме того, время распространения сигнала считается пренебрежимо малым.

Для описания изменений состояния канала используем следующую модель Джилберта: канал пребывает в состоянии i ($i = 0, 1$) в течение случайного интервала, распределенного экспоненциально с параметром λ_i . В состоянии i BER равно $\mu_i^h/8$ и $\mu_i/8$ при передаче соответственно h -байтного заголовка физического уровня и остальной части кадра, т.е. $h + f$ -байтный кадр, передаваемый за интервал времени, когда канал находился только в состоянии i , искажается с вероятностью $1 - \exp\{-\mu_i^h h - \mu_i f\}$. Мы вынуждены использовать разные значения BER, так как заголовок физического уровня обычно передается на более низкой скорости, но и с помощью более надежной схемы кодирования и модуляции. Предполагается, что интенсивности изменения состояния канала λ_0 и λ_1 не слишком велики, так что в течение передачи кадра или межкадрового интервала может произойти не более одного изменения состояния.

Как и в предыдущем разделе, будем оперировать понятием виртуального слота и считать вероятность начала передачи данной станции в произвольно выбранном слоте не зависимой ни от предыстории, ни от поведения остальных станций и равной τ_i (т.е. эта вероятность зависит только от текущего состояния канала i). Следовательно, вероятности «пустого», «успешного» и «коллизийного» слотов, начинающихся, когда канал находится в состоянии i , равны

$$p_e^i = (1 - \tau_i)^N, \quad p_s^i = N\tau_i(1 - \tau_i)^{N-1}, \quad p_c^i = 1 - p_e^i - p_s^i. \quad (6.72)$$

Назовем попытку передачи пакета, происходящую после пребывания передающей станции в состоянии отсрочки, (i, ℓ, k) -попыткой, если эта попытка начинается, когда канал находится в состоянии i , и относится к пакету длиной ℓ байт, k фрагментов которого осталось передать. Пусть $d_{\ell k}^i$ — вероятность того, что произвольно выбранная попытка передачи пакета, начинающаяся при состоянии i , является (i, ℓ, k) -попыткой.

Тогда пропускная способность S в условиях насыщения сети равна

$$S = \sum_{i=0}^1 \psi_i S_i, \quad (6.73)$$

где S_i — пропускная способность, наблюдаемая при состоянии канала i , и $\psi_i = \lambda_{i^*}/(\lambda_0 + \lambda_1)$ — доля времени, когда канал находится в состоянии i (здесь и далее в этом разделе $i^* = 1$ при $i = 0$ и $i^* = 0$ при $i = 1$.)

Аналогично (6.41) имеем

$$S_i = \frac{p_s^i}{T_{sl}^i} \sum_{\ell=1}^L \sum_{k=1}^{K(\ell)} 8\ell Q_{i0}(\ell, k) d_{\ell k}^i, \quad (6.74)$$

где $T_{sl}^i = p_e^i \sigma + p_s^i T_s^i + p_c^i T_c^i$, T_s^i и T_c^i — средние длительности соответственно виртуального, «успешного» и «коллизийного» слотов, начинающихся, когда канал находится в состоянии i . $K(\ell)$ — число фрагментов, на которые разделяется пакет длиной ℓ , т.е. $K(\ell)$ — минимальное целое число, не меньшее ℓ/L_f . $Q_{i0}(\ell, k)$ — вероятность того, что (i, ℓ, k) -попытка, выполняющаяся в «успешный» слот, успешно завершает передачу всего пакета.

Средняя длительность «коллизийного» слота находится аналогично (6.42):

$$T_c^i = \text{EIFS} + (p_c^i)^{-1} \sum_{r=1}^{L_f} t_d(r) \{ \alpha_i^N(r) - \alpha_i^N(r-1) - p_s^i [D(r, i) - D(r-1, i)] \}, \quad (6.75)$$

где $\alpha_i^N(r) = \{1 - \tau_i[1 - D(r, i)]\}^N$, а $D(r, i)$ — вероятность того, что при текущей попытке передается цепочка кадров DATA, первый из которых содержит фрагмент длиной $r_0 \leq r$, т. е. $D(L_f, i) = 1$ и

$$D(r, i) = \sum_{h=1}^r \sum_{k=0}^{k_{\max}(r)} d_{kL_f+r, 1}^i$$

при $r < L_f$ [здесь $k_{\max}(r)$ — целая часть отношения $(\ell_{\max} - r)/L_f$].

Случай «успешного» слота более сложен, и мы рассмотрим его в следующем подразделе.

6.12.2 Успешный слот

Итак, рассмотрим «успешный» слот, начинающийся при состоянии канала i . В его начале только одна станция делает попытку передачи, причем эта попытка — (i, ℓ, k) -попытка с вероятностью $d_{\ell k}^i$, и она успешна (т. е. завершает передачу всего пакета длиной ℓ) с вероятностью $Q_{i0}(\ell, k)$, если ни один из кадров, передаваемых источником и получателем в течение этого процесса, не искажен помехами. Таким образом,

$$T_s^i = \sum_{\ell=1}^L \sum_{k=1}^{K(\ell)} d_{\ell k}^i T_i(\ell, k) Q_{i0}(\ell, k) + \sum_{\ell=1}^L \sum_{k=1}^{K(\ell)} d_{\ell k}^i \sum_{m=1}^k T_i^e(\ell, k, m) [Q_{ii}^m(\ell, k) + Q_{ii^*}^m(\ell, k)], \quad (6.76)$$

где

$$T_i(\ell, k) = (k-1)[t_d(L_f) + \delta] + kt_{ACK} + t_d(r_{\ell}^0) + \delta_d$$

— средняя длительность успешной (i, ℓ, k) -попытки.

$$T_i^e(\ell, k, m) = (m-1)[t_d(L_f) + t_{ACK} + \delta] + t_d(r_{\ell k}^m) + (t_{ACK} + \delta)\Theta_i^a(\ell, k, m) + \delta_e$$

— средняя длительность «успешного» слота при условии, что первые $m-1$ ($m < k$) фрагментов остатка пакета переданы успешно, а фрагмент m длиной $r_{\ell k}^m$ — неудачно. $Q_{ii^*}^m(\ell, k)$ и $Q_{ii}^m(\ell, k)$ — вероятности этого условия в случаях, когда состояния канала в начале и конце этого слота соответственно отличаются и совпадают [здесь и далее в этом подразделе мы находим вероятности отдельно для случаев изменения (индекс « ii^* ») и неизменности (индекс « ii ») состояния канала]. Наконец, $\Theta_i^a(\ell, k, m)$ — вероятность того, что сбой происходит именно из-за искажения последнего кадра АСК. Найдем эти вероятности.

Передача кадра длиной f байт (включая h_{MAC} -байтовый заголовок MAC-уровня), начинающаяся при состоянии канала i , завершается успешно с вероятностями

$$\nu_{ii}^0(f) = \exp\{-\lambda_i(H + f/V) - \mu_i^h h - \mu_i f\}$$

и

$$\nu_{ii^*}^0(f) = e^{-\mu_i^* f} I_i^h + e^{-\mu_i^h h - \lambda_i H} I_i(f),$$

где

$$I_i^h = \int_0^H \lambda_i e^{-\lambda_i t} \exp\{-V_h[\mu_i^h t + \mu_i^*(H-t)]\} dt,$$

т. е.

$$I_i^h = e^{-\mu_i^h h} \frac{\lambda_i H}{\lambda_i H + h(\mu_i^h - \mu_{i^*}^h)} [1 - e^{-\lambda_i H - h(\mu_i^h - \mu_{i^*}^h)}]$$

при $\lambda_i H \neq h(\mu_i^h - \mu_{i^*}^h)$, а иначе $I_i^h = \lambda_i H e^{-\mu_i^h h}$. $I_i(f)$ определяется аналогично с подстановкой f вместо h , f/V вместо H , а также μ_i и μ_{i^*} вместо μ_i^h и $\mu_{i^*}^h$. Передача неудачна с вероятностями

$$\nu_{ii}^1(f) = \exp\{-\lambda_i(H + f/V)\} [1 - \exp\{-\mu_i^h h - \mu_i f\}]$$

и

$$\nu_{ii^*}^1(f) = 1 - \nu_{ii}^0(f) - \nu_{ii^*}^0(f) - \nu_{ii}^1(f).$$

Теперь рассмотрим процесс передачи фрагмента длиной τ , включая возможный прием ответного кадра АСК. Пусть этот процесс стартует при состоянии канала i . Тогда он успешен с вероятностями

$$\eta_{ii}^0(\tau) = \nu_{ii}^0(f_\tau) [\gamma_i \nu_{ii}^0(l_{АСК}) + \bar{\gamma}_i \nu_{i^*i}^0(l_{АСК})] + \nu_{ii^*}^0(f_\tau) [\gamma_{i^*} \nu_{i^*i}^0(l_{АСК}) + \bar{\gamma}_{i^*} \nu_{ii}^0(l_{АСК})]$$

и

$$\begin{aligned} \eta_{ii^*}^0(\tau) = & \nu_{ii^*}^0(f_\tau) [\gamma_{i^*} \nu_{i^*i^*}^0(l_{АСК}) + \bar{\gamma}_{i^*} \nu_{ii}^0(l_{АСК})] + \\ & + \nu_{ii}^0(f_\tau) [\gamma_i \nu_{ii^*}^0(l_{АСК}) + \bar{\gamma}_i \nu_{i^*i}^0(l_{АСК})], \end{aligned}$$

где $f_\tau = \tau + h_{MAC}$, $\gamma_i = 1 - \bar{\gamma}_i = \exp\{-\lambda_i \delta\}$, а $l_{АСК}$ — длина кадра АСК. Этот процесс неудачен с вероятностями

$$\begin{aligned} \eta_{ii}^1(\tau) = & \nu_{ii}^1(f_\tau) + \nu_{ii}^0(f_\tau) [\gamma_i \nu_{ii}^1(l_{АСК}) + \bar{\gamma}_i \nu_{i^*i}^1(l_{АСК})] \\ \eta_{ii^*}^1(\tau) = & 1 - \eta_{ii}^0(\tau) - \eta_{ii^*}^0(\tau) - \eta_{ii}^1(\tau). \end{aligned}$$

Заметим, что этот процесс неудачен из-за искажения кадра DATA с вероятностями $\nu_{ii}^1(f_\tau)$ (при неизменности состояния канала i) и $\nu_{ii^*}^1(f_\tau)$ (при переходе из состояния i в i^*).

Теперь мы можем найти искомые вероятности Q_{ii}^m , $Q_{ii^*}^m$ и Θ_i^a :

$$\begin{aligned} Q_{ii}^m(\ell, k) = & \beta_{ii}^{(m-1)} [\eta_{ii}^1(r_{\ell k}^m) \gamma_i^e + \eta_{ii^*}^1(r_{\ell k}^m) \bar{\gamma}_{i^*}^e] + \beta_{ii^*}^{(m-1)} [\eta_{i^*i^*}^1(r_{\ell k}^m) \bar{\gamma}_{i^*}^e + \eta_{i^*i}^1(r_{\ell k}^m) \gamma_i^e], \\ Q_{ii^*}^m(\ell, k) = & \beta_{ii}^{(m-1)} [\eta_{ii}^1(r_{\ell k}^m) \bar{\gamma}_i^e + \eta_{ii^*}^1(r_{\ell k}^m) \gamma_{i^*}^e] + \beta_{ii^*}^{(m-1)} [\eta_{i^*i^*}^1(r_{\ell k}^m) \gamma_{i^*}^e + \eta_{i^*i}^1(r_{\ell k}^m) \bar{\gamma}_i^e], \\ \Theta_i^a(\ell, k, m) = & 1 - \frac{\Theta_{ii}^m(\ell, k) + \Theta_{ii^*}^m(\ell, k)}{Q_{ii}^m(\ell, k) + Q_{ii^*}^m(\ell, k)}, \end{aligned}$$

где $\gamma_i^e = 1 - \bar{\gamma}_i^e = \exp\{-\lambda_i \text{EIFS}\}$, $\|\beta_{ij}^{(m)}\| = \|\beta_{ij}\|^m$, $i, j = 0, 1$,

$$\beta_{ii} = \eta_{ii}^0(L_f) \gamma_i + \eta_{ii^*}^0(L_f) \bar{\gamma}_{i^*} \quad \text{и} \quad \beta_{ii^*} = \eta_{ii^*}^0(L_f) \gamma_{i^*} + \eta_{ii}^0(L_f) \bar{\gamma}_i.$$

Кроме того, Θ_{ii}^m и $\Theta_{ii^*}^m$ определяются аналогично Q_{ii}^m и $Q_{ii^*}^m$, подставляя соответствующие функции ν^1 с аргументом $r_{\ell k}^m + h_{MAC}$ вместо всех функций ν^1 .

Наконец, вероятность успешной (i, ℓ, k) -попытки, завершающей передачу всего пакета, очевидно равна

$$Q_{i0}(\ell, k) = 1 - \sum_{m=1}^k [Q_{ii}^m(\ell, k) + Q_{ii^*}^m(\ell, k)],$$

причем после этой попытки канал оказывается в состоянии i или i^* с вероятностями

$$Q_{i0}^i(\ell, k) = \beta_{ii}^{(k-1)}[\eta_{ii}^0(r_\ell^0)\gamma_i^d + \eta_{ii^*}^0(r_\ell^0)\bar{\gamma}_{i^*}^d] + \beta_{ii^*}^{(k-1)}[\eta_{ii^*}^0(r_\ell^0)\bar{\gamma}_{i^*}^d + \eta_{ii}^0(r_\ell^0)\gamma_i^d]$$

и

$$Q_{i0}^{i^*}(\ell, k) = Q_{i0}(\ell, k) - Q_{i0}^i(\ell, k),$$

где $\gamma_i^d = 1 - \bar{\gamma}_i^d = \exp\{-\lambda_i \text{DIFS}\}$.

Таким образом, найдены все компоненты формулы (6.76), и пропускная способность S может быть вычислена по формулам (6.74)–(6.76), если известны вероятности начала передачи τ_0 и τ_1 и вероятностное распределение попыток $\{d_{\ell k}^i\}$.

6.12.3 Алгоритм вычисления вероятностей

Здесь, как и в подразделе 6.10.2, мы исследуем процесс передачи пакета длиной ℓ некоторой станцией. Пусть f_ℓ^i и w_ℓ^i — средние количества соответственно попыток передачи и виртуальных слотов, в которые данная станция не передает, в течение этого процесса, причем эти количества подсчитываются, только если канал находился в состоянии i в начале соответствующего виртуального слота. Тогда

$$\tau_i = \sum_{\ell=1}^{\ell_{\max}} d_\ell f_\ell^i / \sum_{\ell=1}^{\ell_{\max}} d_\ell (f_\ell^i + w_\ell^i). \quad (6.77)$$

Для нахождения распределения $\{d_{\ell k}^i\}$, $k = 1, \dots, K(\ell)$, можно использовать следующую формулу:

$$d_{\ell k}^i = d_\ell f_{\ell k}^i / \sum_{u=1}^L \sum_{v=1}^{K(u)} d_u f_{uv}^i, \quad (6.78)$$

где $f_{\ell k}^i$ — среднее число (i, ℓ, k) -попыток, выполняемых при передаче пакета длиной ℓ .

Кроме того, будем искать также усредненную вероятность отказа, определяемую формулой (6.47).

Начнем с поиска f_ℓ^i . Эту величину можно найти по формуле:

$$f_\ell^i = \sum_{i_s=0}^1 \psi_{i_s}^p F_\ell^i[i_s, K(\ell), 0, 0], \quad (6.79)$$

где $\psi_{i_s}^p$ — вероятность того, что процесс передачи пакета начнется при состоянии канала i_s , и легко показать, что

$$\psi_i^p = (1 - \Phi_{i^*}) / (2 - \Phi_0 - \Phi_1), \quad (6.80)$$

где Φ_i — вероятность того, что канал оказывается в том же состоянии i в конце передачи, что и в ее начале.

Функция $F_\ell^i(i_s, k, n_r, n_s)$ представляет собой среднее число попыток, которые будут сделаны до завершения передачи пакета, при следующих условиях:

- остается передать k фрагментов;
- станция только что перешла в состояние отсрочки с конкурентным окном $W(n_r)$;

- в этот момент канал находится в состоянии i_s ;
- значение счетчика n_s равно n_s .

Очевидно, эта функция вычисляется рекурсивно:

$$F_\ell^i(i_s, k, n_r, n_s) = \sum_{j=0}^1 q_j[n_r, i_s] \left\{ f_0^{ij} + \mathbf{1}(n_s < N_s - 1) \sum_{u=0}^1 [p_j^{cc} Q_{ju}^c(t_{\ell k}^{cj}) + \right. \\ \left. + \bar{p}_j^{cc} Q_{ju}^1(\ell, k)] F_\ell^i(u, k, n_r^*, n_s + 1) + \right. \\ \left. \mathbf{1}(k > 1) \sum_{m=2}^k \bar{p}_j^{cc} Q_{ju}^m(\ell, k) F_\ell^i(u, k - m + 1, n_r^*, 1) \right\}, \quad (6.81)$$

где $n_r^* = n_r + 1$ при $n_r < N_r$, а иначе $n_r^* = N_r$.

Кроме того, в (6.81) $f_0^{ij} = \mathbf{1}(i = j)$, $p_j^{cc} = 1 - \bar{p}_j^{cc} = 1 - (1 - \tau_j)^{N-1}$ — вероятность неудачи текущей попытки из-за коллизии, в течение которой канал переходит из состояния j в u с вероятностью $Q_{ju}^c(t_{\ell k}^{cj}) = 1 - Q_{jj}^c(t_{\ell k}^{cj})$, где

$$Q_{jj}^c(t) = \bar{\gamma}_u^e + (1 - \bar{\gamma}_u^e / \gamma_j^e) \exp\{-\lambda_j t\},$$

а $t_{\ell k}^{cj}$ — средняя длительность такой коллизии, определяемая аналогично (6.75):

$$t_{\ell k}^{cj} = \text{EIFS} + (p_j^{cc})^{-1} \sum_{m=1}^{N-1} \binom{N-1}{m} \tau_i^m (1 - \tau_i)^{N-1-m} \hat{t}_{\ell k}^{cjm},$$

где

$$\hat{t}_{\ell k}^{cjm} = t_d(r_{\ell k}^1) [D(r_{\ell k}^1, j)]^m + \sum_{r=r_{\ell k}^1+1}^{L_f} t_d(r) \{ [D(r, j)]^m - [D(r-1, j)]^m \}$$

— средняя длительность такой коллизии при условии, что в ней участвуют ровно m из остальных $N - 1$ станций. После преобразований получаем:

$$t_{\ell k}^{cj} = \text{EIFS} + \frac{t_d(r_{\ell k}^1)}{p_j^{cc}} [\alpha_j^{N-1}(r_{\ell k}^1) - \bar{p}_j^{cc}] + (p_j^{cc})^{-1} \sum_{r=r_{\ell k}^1+1}^{L_f} t_d(r) [\alpha_j^{N-1}(r) - \alpha_j^{N-1}(r-1)].$$

Наконец, $q_j[n_r, i_s]$ — вероятность того, что канал окажется в состоянии j по завершении времени отсрочки рассматриваемой станции, чье текущее конкурентное окно равно $W(n_r)$, и эта вероятность определяется при условии, что канал был в состоянии i_s в начале отсрочки. Легко показать, что

$$q_j[n_r, i_s] = \frac{1}{W(n_r)} \sum_{b=0}^{W(n_r)-1} p_{i_s j}^{(b)},$$

где $\|p_{ij}^{(b)}\| = \|p_{ij}\|^b$, $i, j = 0, 1$, а p_{ij} — вероятность перехода канала из состояния i в j за виртуальный слот, в течение которого данная станция не передавала. Очевидно,

$$p_{ii} = 1 - p_{ij} = \hat{p}_e^i e^{-\lambda_i \sigma} + \hat{p}_c^i Q_{ii}^c(\hat{T}_c^i) + \hat{p}_s^i Q_{is}^{ii},$$

где $\hat{p}_e^i$, $\hat{p}_s^i$, $\hat{p}_c^i$ и $\hat{T}_c^i$ определяются аналогично p_e^i , p_s^i , p_c^i и T_c^i с подстановкой $N-1$ вместо N . Кроме того,

$$Q_s^{ii} = \sum_{\ell=1}^{\ell_{\max}} \sum_{k=1}^{K(\ell)} d_{\ell k}^i \left[Q_{i0}^i(\ell, k) + \sum_{m=1}^k Q_{ii}^m(\ell, k) \right]$$

— вероятность того, что в конце «успешного» слота канал будет в том же состоянии i , что и в его начале.

Можно применить формулу вида (6.79) и для нахождения w_{ℓ}^i , $f_{\ell v}^i$ и $p_{rej}(\ell)$, используя соответственно функции

$$\bar{W}_{\ell}^i[i_s, K(\ell), 0, 0], \quad \bar{F}_{\ell v}^i[i_s, K(\ell), 0, 0] \quad \text{и} \quad R_{\ell}[i_s, K(\ell), 0, 0]$$

вместо $F_{\ell}^i(\cdot)$, в то время как Φ_i определяется формулой

$$\Phi_i = \sum_{\ell=1}^{\ell_{\max}} d_{\ell} \phi_{\ell}^i[i, K(\ell), 0, 0]. \quad (6.82)$$

Функции $\bar{W}_{\ell}^i(\cdot)$, $\bar{F}_{\ell v}^i(\cdot)$, $R_{\ell}(\cdot)$ и $\phi_{\ell}^i(\cdot)$ с аргументами i_s , k , n_r и n_s в свою очередь, определяются с помощью формулы (6.81), преобразованной следующим образом.

- Для $\bar{W}_{\ell}^i(\cdot)$ исключается f_0^{ij} , а правая часть (6.81) начинается со слагаемого

$$w_{i_s}^i(n_r) = \sum_{b=1}^{W(n_r)-1} \frac{W(n_r) - b}{W(n_r)} p_{i_s i}^{(b-1)}.$$

- Для $\bar{F}_{\ell v}^i(\cdot)$ подставляем $f_1^{ij}(k, v) = 1(i=j)1(k=v)$ вместо f_0^{ij} и оператор $1(k > v)$ вместо $1(k > 1)$.

- Для $R_{\ell}(\cdot)$ заменяем f_0^{ij} на

$$\tau_0^j(\ell, k, n_s) = 1(n_s = N_s - 1) \{ p_j^{cc} + \bar{p}_j^{cc} [\eta_{jj}^1(r_{\ell k}^1) + \eta_{jj}^1(r_{\ell k}^1)] \}.$$

- Для $\phi_{\ell}^i(\cdot)$ f_0^{ij} заменяется на

$$\phi_0^{ij}(\ell, k, n_s) = \bar{p}_j^{cc} Q_{j0}^i(\ell, k) + 1(n_s = N_s - 1) \{ p_j^{cc} Q_{ji}^c(t_{\ell k}^{cj}) + \bar{p}_j^{cc} Q_{ji}^1(\ell, k) \}.$$

Таким образом, пропускную способность S и усредненную вероятность отказа можно вычислить с помощью следующей итеративной процедуры. Сначала вычисляем $Q_{ij}^m(\ell, k)$ и $Q_{i0}^j(\ell, k)$ для всех возможных i , j , ℓ , $k = 1, \dots, K(\ell)$ и $m = 1, \dots, k$. Затем определяем начальные значения τ_i , Ψ_i^p , S , $\bar{p}_{rej}$ и $d_{\ell k}^i$ при всех возможных i , ℓ , k и вычисляем их модифицированные значения по формулам (6.73)–(6.81). Если отличия начальных и модифицированных значений S и $\bar{p}_{rej}$ больше малого, заранее определенного порога, устанавливаем новые начальные значения τ_i , Ψ_i^p , S , $\bar{p}_{rej}$ и $d_{\ell k}^i$ равными полусумме их модифицированных и старых начальных значений. Эта процедура доказала свою быструю сходимость к решению на практике, при оценке производительности БЛС с разными параметрами конфигурации и при различных параметрах модели помех.

6.12.4 Численные результаты

Для оценки точности проведено сравнение результатов, полученных с помощью данной аналитической модели и имитационного моделирования на языке GPSS [157, 195]. Объектом численных исследований являлась БЛС архитектуры ad hoc, работающая под управлением протокола IEEE 802.11b в режиме насыщения. Значения параметров протокола, использованные при моделировании и отличающиеся от указанных в табл. 6.4, приведены в табл. 6.6. Кроме того, размер пакета ℓ (в байтах) выбирался равновероятно из множества $\{1, \dots, 2000\}$, и применялось следующее соотношение между значениями BER при передаче РНУ- и МАС-частей кадров: $\mu_i^h = 0,1\mu_i$, $i = 0, 1$.

Таблица 6.6. Значения параметров протокола (длинная преамбула)

РНУ-заголовок, h	24 байт
МАС-заголовок, $h_{\text{МАС}}$	34 байт
Время передачи РНУ-заголовка, H	192 мкс
Время передачи МАС-заголовка, $H_{\text{МАС}}$	25 мкс
Длина кадра АСК, $l_{\text{АСК}}$	14 байт
Время передачи кадра АСК, $t_{\text{АСК}}$	202 мкс
EIFS	364 мкс

На рис. 6.34 и 6.35 представлены некоторые результаты исследования пропускной способности и вероятности отказа при варьировании среднего BER и фиксированном $N = 20$. Рисунки 6.34, б и 6.35, б соответствуют случаю независимых сбоев канала, а кривые на рис. 6.34, а и 6.35, а получены для случая коррелированных сбоев со следующими параметрами модели канала: $\lambda_0 = 10^{-5}$, $\lambda_1 = 5\lambda_0$, $\mu_0 = 4 \cdot \text{BER}$, $\mu_1 = 28 \cdot \text{BER}$ (состояние 0 считается «хорошим»). Здесь и далее в этом подразделе кривые n/f , f и $o.f$ соответствуют передаче без фрагментации, с фиксированной фрагментацией, т. е. с фиксированным порогом $L_f = 560$ байт, и с оптимальной фрагментацией, когда порог выбирается (с помощью описанной аналитической модели) оптимально в зависимости от значений параметров канала. Критерий оптимизации — максимальная пропускная способность. Сравнение кривых n/f и f , полученных с помощью как аналитической (сплошные линии), так и имитационной (пунктирные линии) моделей, показывает высокую точность аналитического метода: погрешность не превышает 3% при оценке пропускной способности и 6% при оценке вероятности отказа.

Рисунки 6.34 и 6.35 показывают, что как пропускная способность, так и вероятность отказа намного чувствительней к росту среднего BER при независимых сбоях, чем при коррелированных, особенно в случае нефрагментированной передаче: при очень высоком среднем $\text{BER} = 1,5 \times 10^{-4}$ имеем $S = 1,1$ Мбит/с и $\bar{p}_{rej} = 0,19$ для независимых сбоев, а для коррелированных — $S = 2,0$ Мбит/с и $\bar{p}_{rej} = 0,07$. Фрагментация пакетов позволяет снизить влияние помех и сближает значения показателей производительности, показанные при независимых и коррелированных сбоях: для среднего $\text{BER} = 1,5 \times 10^{-4}$ и фиксированной фрагментации имеем $S = 1,9$ Мбит/с и $\bar{p}_{rej} = 0,033$ при независимых сбоях и

$S = 2,3$ Мбит/с и $\bar{p}_{rej} = 0,029$ при коррелированных сбоях. Оптимизация порога фрагментации позволяет лишь слегка (не более чем на 10% в показанных случаях) улучшить пропускную способность по сравнению с максимумом из пары значений, полученных при фиксированной фрагментации и нефрагментированной передаче.

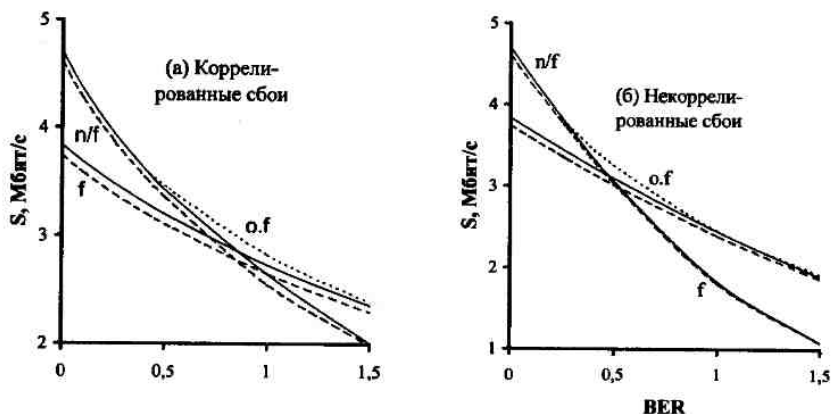


Рис. 6.34. Зависимость пропускной способности S от $BER \cdot 10^4$

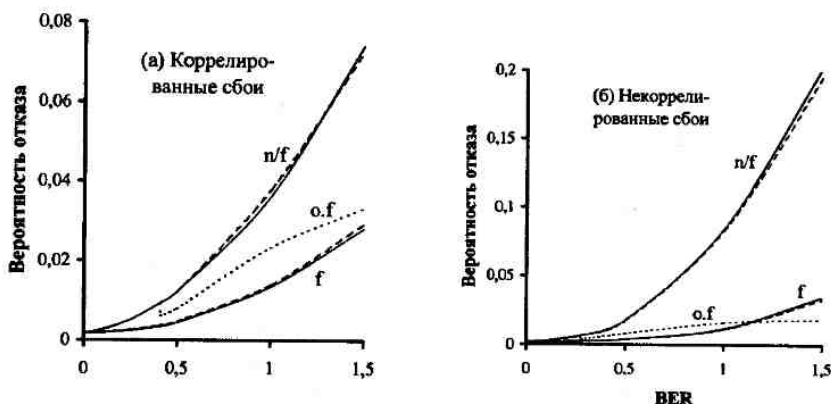


Рис. 6.35. Зависимость вероятности отказа $\bar{p}_{rej}$ от $BER \cdot 10^4$

Теперь исследуем влияние параметров корреляции на пропускную способность и эффективность фрагментации. На рис. 6.36 показаны зависимости пропускной способности от вероятности искажения байта μ_0 , наблюдаемой при «хорошем» состоянии канала, для случая, когда среднее значение BER равно 10^{-4} , $\lambda_0 = 10^{-5}$ и $N = 5$. Семейства кривых (а) и (б) соответствуют: (а) $\lambda_1 = \lambda_0$ и (б) $\lambda_1 = 5\lambda_0$, причем в каждом из этих семейств сплошные, штриховые и пунктирные линии относятся соответственно к нефрагментированной передаче, фиксированной и оптимальной фрагментации. Сравнивая рассматриваемые случаи, можно видеть, что чем меньше BER в «хорошем» состоянии и средняя длитель-

ность пребывания канала в «плохом» состоянии, тем выше пропускная способность и тем менее эффективна фрагментация.

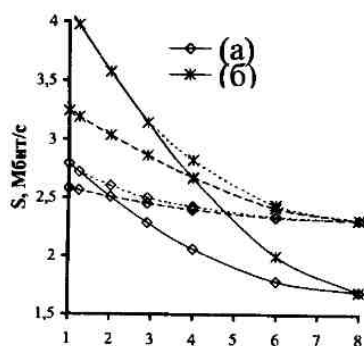


Рис. 6.36. Зависимости пропускной способности от вероятности искажения байта $\mu_0 \cdot 10^4$ при «хорошем» состоянии канала

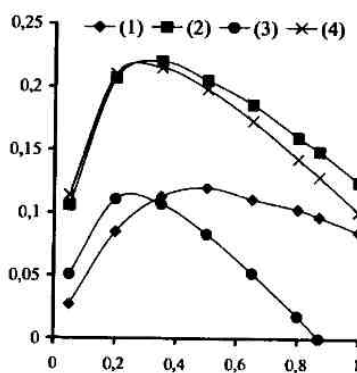


Рис. 6.37. Области эффективности фрагментации (абсцисса — λ_0/λ_1 , ордината — μ_0/μ_1)

На рис. 6.37 показаны области эффективности фрагментации в следующих случаях:

- (1) $\text{BER} = 10^{-4}$, $N = 5$, $\lambda_0 = 10^{-5}$;
- (2) $\text{BER} = 5 \times 10^{-5}$, $N = 5$, $\lambda_0 = 10^{-5}$;
- (3) $\text{BER} = 5 \times 10^{-5}$, $N = 20$, $\lambda_0 = 10^{-5}$;
- (4) $\text{BER} = 5 \times 10^{-5}$, $N = 5$, $\lambda_0 = 2 \times 10^{-6}$.

Для каждого из этих случаев точки, расположенные ниже соответствующей кривой, образуют область эффективности фрагментации, где фрагментация с оптимально выбранным порогом обеспечивает большую пропускную способность, чем нефрагментированная передача. Прежде всего заметим, что ограничивающие кривые немонотонны: эффективность фрагментации высока как в случае примерного равенства средних длительностей пребывания канала в каждом из состояний, так и в случае, когда канал находится в основном в «хорошем» состоянии. Как можно было ожидать, область эффективности фрагментации расширяется с ростом среднего BER [кривые (1) и (2)]. Далее, видно [кривые (2) и (3)], что эта область также расширяется с ростом числа N активных станций. Наконец, следует отметить, что совместный рост длительностей пребывания канала в каждом из состояний [кривые (2) и (4)] не влияет сильно на область эффективности фрагментации.

6.13 Производительность локальных сетей при нормальной нагрузке

В предшествующих разделах, как и в подавляющем большинстве аналитических работ (см., например, [130, 137]), методы оценки пропускной способности БЛС, управляемых схемой 802.11 DCF, разработаны только для высокой нагрузки (случая насыщения), когда в любой момент времени ко всем станциям

БЛС всегда имеются непустые очереди. К сожалению, полученные результаты оказываются неприменимыми в условиях нормальной нагрузки, когда очереди станций периодически оказываются пустыми ввиду существенного завышения оценок среднего времени обслуживания пакетов.

Попытка описать БЛС в условиях нормальной нагрузки предпринята в [235]. Однако в модели, применяемой в [235], не учтены два важных фактора:

1) станция после передачи пакета и при отсутствии следующего пакета в очереди также переходит в состояние отсрочки с минимальным конкурентным окном и лишь по истечении времени отсрочки — в состояние простоя (если очередь осталась пуста);

2) пакет, поступивший в очередь станции, находящейся в состоянии простоя, немедленно передается, если как в момент поступления, так и в течение предшествующего интервала DIFS или EIFS канал был свободен.

Основная цель моделирования, проводимого в данном разделе, — найти среднее значение T времени обслуживания пакета, отсчитываемого от момента либо поступления пакета в пустую очередь данной станции, либо окончания обслуживания предыдущего пакета из этой очереди и до момента либо получения подтверждения ACK, либо истечения интервала EIFS после последней неудачной попытки передачи (т. е. в случае потери пакета).

Исследуем БЛС, состоящую из N статистически одинаковых станций, в очередь каждой из которых поступает пуассоновский поток пакетов с интенсивностью λ и распределением $D(l_i)$ длин пакетов l_i . Канал считается идеальным, а время распространения сигнала между станциями — пренебрежимо малым. Кроме того, предполагается, что очередь пакетов каждой станции может содержать не более B пакетов, порог RTS (который в этом разделе обозначен через L) одинаков для всех станций, а время распространения сигнала пренебрежимо мало.

Будем называть пакеты, передача которых начинается в момент поступления, переданными асинхронно, а все остальные — переданными синхронно. Асинхронная передача имеет место, если в момент прихода пакета станция была в состоянии простоя и канал был свободен в течение как минимум DIFS или EIFS. Таким образом, асинхронная передача происходит только при отсутствии синхронных передач других станций, а так как $N\lambda\sigma \ll 1$, то можно считать, что за время одного слота отсрочки в сети может произойти не более одной асинхронной передачи. Следовательно, асинхронная передача всегда успешна.

6.13.1 Модель Маркова

Для оценки времени T построим модель поведения исследуемой станции в виде цепи Маркова с дискретным целочисленным временем, единицей которого является виртуальный слот — промежуток времени между последовательным изменением счетчика отсрочки у каждой станции, не находящейся в состоянии простоя. Пусть $b(t)$ — стохастический процесс изменения счетчика отсрочки для данной станции, моменты времени t и $t+1$ соответствуют началу двух последовательных виртуальных слотов, причем, станция передает, когда $b(t) = 0$. В то же время $s(t)$ — стохастический процесс изменения стадии отсрочки $0, \dots, m$ (номер которой равен числу сделанных попыток передачи текущего пакета, т. е. в

данном разделе $m = N_s$), расширенный (по сравнению с моделью Бьянки [130]) значением -1 для ситуации, когда в очереди нет пакета.

Заметим, что, исходя из принятой модели, эти виртуальные времена не имеют прямого соответствия реальному времени и виртуальные слоты неоднородны. Как уже было сказано, счетчик отсрочки «замораживается», если станция замечает передачу другой станции. Поэтому реальное время, прошедшее между t и $t + 1$, больше слота отсрочки σ при наличии передачи другой станции. Таким образом, как и ранее, имеем три вида виртуальных слотов: 1) «пустой» слот, во время которого ни одна станция не вела передачу; 2) «успешный» слот, когда одна и только одна станция вела передачу; 3) «коллизийный» слот, во время которого произошла коллизия.

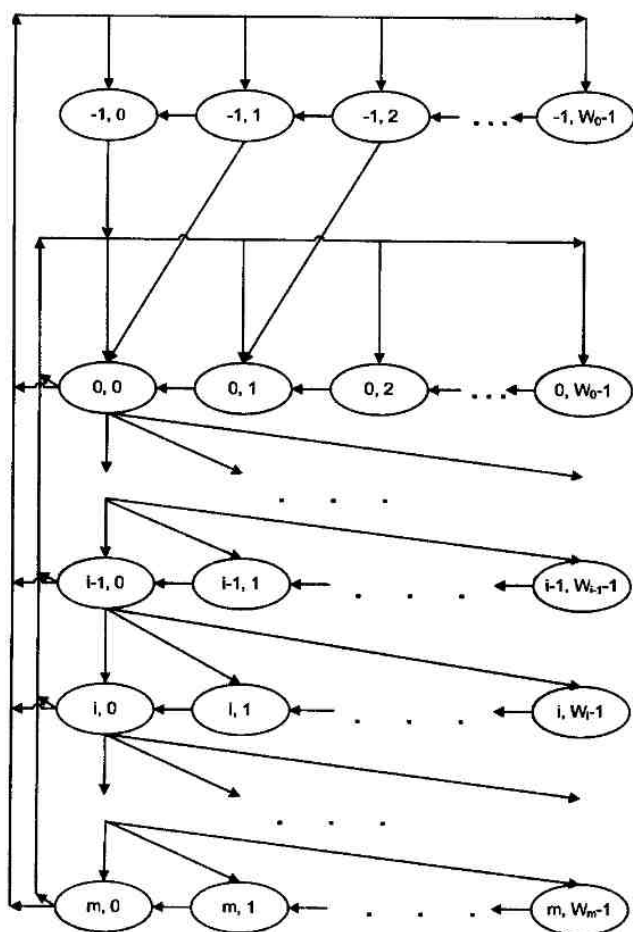


Рис. 6.38. Цепь Маркова

Двумерный процесс $\{s(t), b(t)\}$ опишем цепью Маркова, показанной на рис. 6.38, где состоянию простаивающей станции соответствует состояние $(-1, 0)$. Состояния, когда станция не имеет пакета для передачи, но выполняет процедуру отсрочки после

удачной передачи или отказа, — это $(-1, 1, \dots, W_0 - 1)$. Наконец, состояния, когда станция имеет пакет и выполняет процедуру отсрочки, — это все остальные (i, k) , где $k = 0, \dots, W_i - 1$ — значение счетчика отсрочки, а $i = 0, \dots, m$ — стадия отсрочки.

Пусть $\alpha(i, k)$ — стационарная вероятность состояния (i, k) , а $P\{i_2, k_2 | i_1, k_1\}$ — вероятность одношагового перехода из (i_1, k_1) в (i_2, k_2) . Введем следующие обозначения:

P_0 — вероятность опустошения очереди после завершения синхронного обслуживания;

P_S — вероятность прихода хотя бы одного пакета за время виртуального слота при условии что очередь данной станции пуста. Очевидно, эта вероятность включает в себя два компонента: $P_S = P_S^F + P_S^E$, где P_S^F и P_S^E — вероятности прихода хотя бы одного пакета за время соответственно непустого и пустого слота, при условии, что очередь данной станции пуста;

P_T — вероятность прихода хотя бы одного пакета за время успешной передачи другого пакета;

p — вероятность неудачной попытки передачи из-за коллизии (вероятность коллизии). Как и в [130], считаем, что она не зависит от стадии отсрочки i .

Определим возможные одношаговые переходы между состояниями и соответствующие им ненулевые вероятности переходов:

$P\{i, k | i, k + 1\} = 1, i \in (0, m), k \in (0, W_i - 2)$ — уменьшение счетчика отсрочки;

$P\{i, k | i - 1, 0\} = p/W_i, i \in (1, m), k \in (0, W_i - 1)$ — неудачная попытка передачи и переход на следующую стадию отсрочки;

$P\{0, k | i, 0\} = (1 - P_0 e^{-\lambda \text{DIFS}})(1 - p)/W_0, i \in (0, m - 1), k \in (0, W_0 - 1)$ — удачная передача, в очереди есть еще пакет(ы);

$P\{-1, k | i, 0\} = P_0 e^{-\lambda \text{DIFS}}(1 - p)/W_0, i \in (0, m - 1), k \in (0, W_0 - 1)$ — удачная передача, в очереди нет пакетов;

$P\{0, k | m, 0\} = [(1 - P_0 e^{-\lambda \text{DIFS}})(1 - p) + (1 - P_0 e^{-\lambda \text{EIFS}})p]/W_0, k \in (0, W_0 - 1)$ — последняя попытка передачи, после которой пакет удаляется из очереди; в очереди есть еще пакет(ы);

$P\{-1, k | m, 0\} = P_0[e^{-\lambda \text{DIFS}}(1 - p) + e^{-\lambda \text{EIFS}}p]/W_0, k \in (0, W_0 - 1)$ — последняя попытка передать пакет, в очереди больше нет пакетов;

$P\{0, k | -1, k + 1\} = P_S, k \in (0, W_0 - 2)$ — уменьшение счетчика отсрочки, и в пустую очередь пришел пакет;

$P\{-1, k | -1, k + 1\} = 1 - P_S, k \in (0, W_0 - 2)$ — уменьшение счетчика отсрочки, очередь осталась пуста;

$P\{0, k | -1, 0\} = (P_S^F + P_S^E P_T)/W_0, k \in (0, W_0 - 1)$ — переход из состояния простоя в состояние отсрочки. Такой переход имеет место, если в момент прихода пакета канал был занят или в момент асинхронной передачи пришел еще один пакет;

$P\{-1, k | -1, 0\} = P_S^E(1 - P_T)/W_0, k \in (1, W_0 - 1)$ — переход соответствует асинхронной передаче, после которой в очереди нет больше пакетов и начальное значение счетчика отсрочки $b = k > 0$;

$P\{-1, 0 | -1, 0\} = 1 - P_S + P_S^E(1 - P_T)/W_0$ — нет поступивших пакетов или имела место асинхронная передача, за время которой не поступило больше пакетов и $b = 0$.

Очевидно,

$$\sum_{i=0}^m \sum_{k=0}^{W_i-1} \alpha(i, k) + \sum_{k=0}^{W_0-1} \alpha(-1, k) = 1. \quad (6.83)$$

Для $i \in (1, m)$ и $k \in (0, W_i - 1)$, т.е. состояний, соответствующих процедуре отсрочки при имеющемся пакете для передачи и уже не первой попытке передать его, стационарные вероятности определяются формулами:

$$\begin{aligned} \alpha(i, k) &= \frac{p}{W_i} \alpha(i-1, 0) + \alpha(i, k+1), \\ \alpha(i, 0) &= \frac{p}{W_i} \alpha(-1, 0) + \alpha(i, 1) = \dots = p \alpha(-1, 0), \end{aligned}$$

т.е.

$$\alpha(i, k) = \frac{W_i - k}{W_i} p \alpha(i-1, 0), \quad \alpha(i, 0) = p^i \alpha(0, 0),$$

и их сумма имеет вид

$$\sum_{i=1}^m \sum_{k=0}^{W_i-1} \alpha(i, k) = \sum_{i=1}^m \frac{W_i + 1}{2} p^i \alpha(0, 0). \quad (6.84)$$

Для $i = -1$ и $k \in (1, W_0 - 1)$, т.е. состояний, соответствующих процедуре отсрочки после удачно переданного пакета или отказа, но при отсутствии пакета для передачи, из уравнения глобального баланса имеем:

$$\alpha(-1, k) = \alpha(0, 0) P_S \widehat{P}_0 A(k) / C,$$

а для состояния простоя $(-1, 0)$:

$$\alpha(-1, 0) = \alpha(0, 0) \widehat{P}_0 A / C,$$

где

$$\begin{aligned} \widehat{P}_0 &= P_0 [e^{-\lambda \text{DIFS}} (1 - p^{m+1}) + e^{-\lambda \text{EIFS}} p^{m+1}], \\ A(k) &= \sum_{t=k}^{W_0-1} (1 - P_S)^{W_0-1-t}, \quad k = 0, \dots, W_0 - 1, \quad A = A(0), \\ C &= P_S W_0 + P_S^E (1 - P_T) A. \end{aligned}$$

После сложения и простых преобразований получаем:

$$\sum_{k=0}^{W_0-1} \alpha(-1, k) = \alpha(0, 0) \widehat{P}_0 W_0 / C. \quad (6.85)$$

Для $i = 0$ и $k = 1, \dots, W_0 - 1$ после преобразований имеем:

$$\alpha(0, k) = \left[\frac{W_0 - k}{W_0} \left(1 - \widehat{P}_0 + (P_S^F + P_S^E P_T) \frac{\widehat{P}_0 A}{C} \right) + \frac{P_S^2 \widehat{P}_0}{C} \sum_{t=k+1}^{W_0-1} A(t) \right] \alpha(0, 0). \quad (6.86)$$

Наконец, из (6.83)–(6.86) получаем:

$$\alpha(0,0)^{-1} = 1 + \sum_{i=1}^m \frac{W_i + 1}{2} p^i + \frac{W_0 \widehat{P}_0}{C} + \frac{W_0 - 1}{2} \left(1 - \widehat{P}_0 + (P_S^F + P_S^E P_T) \frac{\widehat{P}_0 A}{C} \right) + \frac{P_S^2 \widehat{P}_0}{C} \sum_{k=1}^{W_0-2} \sum_{t=k+1}^{W_0-1} A(t).$$

Пусть τ — вероятность синхронной передачи данной станции за время виртуального слота. Очевидно, что

$$\tau = \sum_{i=0}^m \alpha(i, 0) = \sum_{i=0}^m p^i \alpha(0, 0) = \frac{1 - p^{m+1}}{1 - p} \alpha(0, 0).$$

Считая независимыми стохастические процессы $\{s(t), b(t)\}$ всех станций, находим вероятность того, что синхронная передача данной станции будет неудачна из-за коллизии:

$$p = 1 - (1 - \tau)^{N-1}.$$

Вероятность же асинхронной передачи за время виртуального слота равна $\tau_a = \alpha(-1, 0) P_S^E$.

Перейдем теперь к определению вероятностей P_S , P_S^F , P_S^E , P_T . Так как согласно принятому допущению за один виртуальный слот может произойти только одна асинхронная передача, вероятность прихода в очередь данной станции хотя бы одного пакета за время пустого слота при условии, что эта очередь пуста, равна $P_S^E = (1 - \tau)^{N-1} (1 - e^{-\lambda \sigma})$.

Для того чтобы определить вероятности P_S^F и P_T , найдем времена «непустых» слотов, за которые произошли либо успешная передача, либо коллизия. Очевидно, время успешной передачи пакета длиной l_i равно: $t_i^S = l_i/V + t_{\text{const}}^S$ при $l_i \leq L$ и $t_i^S = t_{\text{RTS}} + t_{\text{CTS}} + l_i/V + 2\text{SIFS} + t_{\text{const}}^S$ при $l_i > L$, где $t_{\text{const}}^S = H + t_{\text{ACK}} + \text{SIFS} + \text{DIFS}$. Таким образом, вероятность прихода хотя бы одного пакета за время успешной передачи: $P_T = 1 - \sum_i e^{-\lambda t_i^S} D(l_i)$.

При определении распределения времени коллизии используем тот же метод, что и в подразделе 6.10.1, с учетом того, что в начале рассматриваемого виртуального слота данная станция не передает. Аналогично (6.42) получаем следующие формулы. Время коллизии равно $t_{L+1}^C = t_{\text{RTS}} + \text{EIFS}$ с вероятностью

$$D_{L+1}^C = [p_c^{(N-1)}] \left[(1 - \tau + \tau F_0)^{N-1} - p_e^{(N-1)} - F_0 p_s^{(N-1)} \right],$$

где

$$F_0 = \sum_{j: l_j > L} D(l_j),$$

а $p_e^{(N-1)}$, $p_s^{(N-1)}$ и $p_c^{(N-1)}$ определяются аналогично p_e , p_s и p_c в (6.3) с заменой N на $N - 1$. Время коллизии равно $t_i^C = \frac{l_i}{V} + H + \text{EIFS}$ при $l_i \leq L$ с вероятностью

$$D_i^C = [p_c^{(N-1)}] \left[(1 - \tau + \tau F_i)^{N-1} - (1 - \tau + \tau F_{i-1})^{N-1} - D(l_i) p_s^{(N-1)} \right],$$

где

$$F_i = F_0 + \sum_{j: l_j \leq l_i} D(l_j).$$

Таким образом, вероятность прихода хотя бы одного пакета за время коллизии равна

$$P_C = 1 - \sum_{i \leq L+1} e^{-\lambda t_i^C} D_i^C.$$

P_S^F — вероятность прихода пакета за время непустого слота, при условии что очередь данной станции пуста. Рассмотрим три случая, которые могут иметь место.

1. Синхронная успешная передача другой станции. Вероятность поступления пакета в этом случае равна $Q_S^S P_T$, где $Q_S^S = (N-1)\tau(1-\tau)^{N-2}$ — условная вероятность этого случая.

2. Асинхронная передача другой станции. При анализе этого случая используем допущение о том, что за один виртуальный слот может произойти только одна асинхронная передача, и она успешна. Тогда условная вероятность этого случая равна $Q_A = (N-1)\tau_a(1-\tau)^{N-2}$, а вероятность поступления равна $Q_A P_T$.

3. В случае коллизии вероятность поступления равна $Q_S^C P_C$, где $Q_S^C = 1 - Q_E - Q_S^S - Q_A$ — вероятность коллизии, в которой не участвует данная станция.

Следовательно,

$$P_S^F = (Q_S^S + Q_A) P_T + Q_S^C P_C.$$

Для завершения определения модели осталось найти P_0 — вероятность опустошения очереди после завершения обслуживания.

Процесс изменения очереди можно описать следующей моделью, показанной на рис. 6.39.

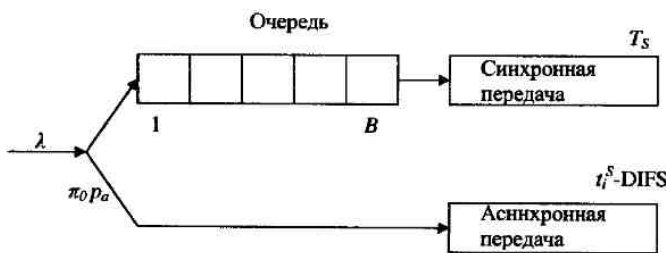


Рис. 6.39. Процесс изменения очереди

Пакеты, приходящие на станцию, не занятую обслуживанием других пакетов, с вероятностью p_a обслуживаются асинхронно и поэтому успешны в течение $t_i^S - \text{DIFS}$ (при длине пакета l_i). В остальных случаях они поступают в буфер размером B и обслуживаются синхронно в течение случайного времени со средним значением T_S .

Предположим, что время синхронного обслуживания распределено экспоненциально. Тогда изменение очереди синхронного обслуживания пакетов, очевидно, описывается процессом рождения-гибели, стационарная вероятность состояния i

которого равна

$$\pi_i = \pi_0 \lambda_0 \lambda^{i-1} T_S^i, \quad i = 1, \dots, B, \quad (6.87)$$

где $\lambda_0 = (1 - p_a) \lambda$. Следовательно, вероятность опустошения очереди после завершения синхронного обслуживания равна $P_0 = \pi_1 / (1 - \pi_0)$, а так как

$$\pi_0 = \frac{1}{1 + (1 - p_a) \sum_{i=1}^B (\lambda T_S)^i}, \quad (6.88)$$

то

$$P_0 = \frac{1}{\sum_{i=1}^B (\lambda T_S)^{i-1}} = \frac{1 - \lambda T_S}{1 - (\lambda T_S)^B}. \quad (6.89)$$

Оценку вероятности p_a , а также времени синхронного обслуживания T_S и основного показателя производительности T проведем в следующем подразделе.

6.13.2 Оценка показателей производительности

Для нахождения вероятности p_a и среднего времени синхронного обслуживания разобьем пакеты, обслуживаемые синхронно и поступающие в течение всех возможных виртуальных слотов (i, k) , на следующие 4 категории:

- 1) поступление в течение слотов $(i \geq 0, k)$;
- 2) поступление в течение слотов $(-1, k > 0)$;
- 3) поступление в течение передачи другой станции во время слота $(-1, 0)$;
- 4) поступление во время асинхронной передачи данной станции.

Для каждой из этих категорий подсчитаем среднее количество поступающих синхронно обслуживаемых пакетов n_i и n_i^0 ($i = 1, \dots, 4$), где n_i — их общее количество, а n_i^0 — количество таких пакетов, поступающих в пустую очередь.

Очевидно,

$$T_S = \left[(T_S^* + \Delta^*) \sum_{i=1}^4 (n_i - n_i^0) + \sum_{i=1}^4 T_S^i n_i^0 \right] / \sum_{i=1}^4 n_i, \quad (6.90)$$

$$p_a = \tau_a / \left(\tau_a + \sum_{i=1}^4 n_i^0 \right), \quad (6.91)$$

где

$$T_S^* = \sum_{i=0}^m p^i \left(\frac{W_i - 1}{2} t_1 + t_2 - (1 - p) \text{DIFS} \right),$$

а

$$t_1 = Q_E \sigma + (Q_S^S + Q_A) \sum_i t_i^S D(l_i) + Q_S^C \sum_{i \leq L+1} t_i^C D_i^C,$$

$$t_2 = (1 - p) \sum_i t_i^S D(l_i) + p \sum_{i \leq L+1} t_i^C D_i^{C1}$$

— средние длительности виртуальных слотов, в которые данная станция соответственно воздерживается от передачи и передает. Кроме того, T_S^i — среднее время обслуживания пакетов разных категорий i , поступающих в пустую очередь,

а $\Delta^* = (1 - p^{m+1})\text{DIFS}$ отражает тот факт, что передача пакета, поступившего в непустую очередь, начинается с интервала DIFS, следующего за подтверждением успешной передачи, кроме случая отказа, когда обслуживание начинается после окончания интервала EIFS. Наконец, $\{D_i^{C1}\}$ — распределение длительности коллизии, в которой участвует данная станция. Это распределение находится аналогично $\{D_i^C\}$: длительность коллизии равна t_{L+1}^C с вероятностью

$$\begin{aligned} D_{L+1}^{C1} &= p^{-1} F_{L+1} \sum_{h=1}^{N-1} \binom{N-1}{h} (\tau F_{L+1})^h (1-\tau)^{N-h} = \\ &= F_{L+1} [p^{-1} (1-\tau + F_{L+1}\tau)^{N-1} - 1] \end{aligned}$$

и равна t_i^C , $i \leq L$, с вероятностью

$$\begin{aligned} D_i^{C1} &= p^{-1} \sum_{h=1}^{N-1} \binom{N-1}{h} \tau^h (1-\tau)^{N-h} (F_i^{h+1} - F_{i-1}^{h+1}) = \\ &= p^{-1} [F_i (1-\tau + F_i\tau)^{N-1} - F_{i-1} (1-\tau + F_{i-1}\tau)^{N-1}] - D(i). \end{aligned}$$

Найдем значения n_i , n_i^0 и T_S^i для каждой из введенных категорий.

Категория 1. Поступление в течение слотов ($i \geq 0, k$). Пакеты поступают в пустую очередь в течение интервала DIFS (или EIFS в случае отказа), следующего за успешной синхронной передачей (или отказом) данной станцией ее предыдущего пакета. Поэтому $n_1^0 = n_{1s}^0 + n_{1c}^0$, где

$$n_{1s}^0 = (1 - e^{-\lambda \cdot \text{DIFS}}) (1-p) P_0 \sum_{i=0}^m \alpha(i, 0) \quad \text{и} \quad n_{1c}^0 = (1 - e^{-\lambda \cdot \text{EIFS}}) p P_0 \alpha(m, 0),$$

а общее число поступивших пакетов этой категории

$$n_1 = \lambda t_1 \sum_{i=0}^m \sum_{k=1}^{W_i-1} \alpha(i, k) + \lambda t_2 \sum_{i=0}^m \alpha(i, 0) \quad \text{и} \quad T_S^1 = T_S^* + \frac{n_{1s}^0}{n_1^0} \frac{\text{DIFS}}{2} + \frac{n_{1c}^0}{n_1^0} \frac{\text{EIFS}}{2}.$$

Категория 2. Поступление в течение слотов ($-1, k > 0$). Так как только первый пакет, поступивший в течение каждого из данных слотов, поступает в пустую очередь, то

$$n_2^0 = Q^* \sum_{k=1}^{W_0-1} \alpha(-1, k), \quad n_2 = \lambda t_1 \sum_{k=1}^{W_0-1} \alpha(-1, k),$$

где

$$Q^* = Q_E (1 - e^{-\lambda \sigma}) + (Q_S^S + Q_A) \sum_i (1 - e^{-\lambda t_i^S}) D(i) + Q_S^C \sum_{i \leq L+1} (1 - e^{-\lambda t_i^C}) D_i^C.$$

В этом случае при поступлении пакета в пустую очередь среднее время до первой попытки передачи сокращается до $(k - \frac{1}{2}) t_1$, и среднее время обслуживания

$$T_S^2 = T_S^* - \frac{Q^* t_1}{n_2^0} \sum_{k=1}^{W_0-1} \left(\frac{W_0}{2} - k \right) \alpha(-1, k).$$

Категория 3. Пакеты в поступают в течение слота $(-1, 0)$, и в это время идет передача другой станции. Для этого случая

$$n_3^0 = \left[(Q_S^S + Q_A) \sum_i (1 - e^{-\lambda t_i^S}) D(l_i) + Q_S^C \sum_{i \leq L+1} (1 - e^{-\lambda t_i^C}) D_i^C \right] \alpha(-1, 0),$$

$$n_3 = \left[(Q_S^S + Q_A) \lambda \sum_i t_i^S D(l_i) + Q_S^C \lambda \sum_{i \leq L+1} t_i^C D_i^C \right] \alpha(-1, 0),$$

и среднее время обслуживания $T_S^3 = \Delta_3 + T_S^*$, где

$$\Delta_3 = \frac{1}{2(1 - Q_E)} \left[Q_S^S \sum_i t_i^S D(l_i) + Q_S^C \sum_{i \leq L+1} t_i^C D_i^C \right].$$

Категория 4. Пакет поступает в течение слота $(-1, 0)$, и в течение этого слота уже происходит асинхронная передача этой станции. Тогда среднее время обслуживания для пакета, поступившего в пустую очередь синхронно обслуживаемых пакетов в течение асинхронной передачи этой же станции, равно $T_S^4 = T_S^* + \frac{1}{2} \sum_i t_i^S D(l_i)$, а количество таких пакетов — $n_4^0 = \tau_a P_T$, в то время как общее количество пакетов этой категории $n_4 = \tau_a \lambda \sum_i t_i^S D(l_i)$.

Окончательно, учитывая асинхронную передачу, получаем среднее время, затрачиваемое на обслуживание пакета (исключая пакеты, «отвергнутые» из-за переполнения буфера):

$$T = \frac{\pi_0 p_a}{1 - \pi_B} \sum_i (t_i^S - \text{DIFS}) D(l_i) + \frac{1 - \pi_0 p_a - \pi_B}{1 - \pi_B} T_S.$$

Здесь первое слагаемое ответственно за асинхронный, а второе — за синхронный механизм передачи.

В заключение найдем другие показатели производительности. Очевидно, вероятность отказа в обслуживании пакета

$$p_r = 1 - (1 - \pi_B) [1 - p^{m+1} (1 - p_a)].$$

Отказ происходит при: а) полном заполнение очереди, когда количество пакетов в ней равно B — вероятность π_B , и б) исчерпаниии количества попыток на передачу пакета — вероятность $p^{m+1} (1 - p_a)$.

Наконец, на основании формулы Литтла находим среднее время задержки пакета на МАС-уровне (также за исключением пакетов, «отвергнутых» из-за переполнения буфера), т.е. среднее время его пребывания на данной станции, включая возможное ожидание в очереди и обслуживание:

$$T_{MAC} = \frac{\pi_0 p_a}{1 - \pi_B} \sum_i t_i^S D(l_i) + \frac{1 - \pi_0 p_a - \pi_B}{1 - \pi_B} \frac{\sum_{i=1}^B i \pi_i}{\lambda_0 \pi_0 + \lambda \sum_{i=1}^{B-1} \pi_i} =$$

$$= \frac{\pi_0 p_a}{1 - \pi_B} \sum_i t_i^S D(l_i) + \frac{1}{(1 - \pi_B) \lambda} \sum_{i=1}^B i \pi_i.$$

6.14 Моделирование широковещательной беспроводной сети

Применим результаты, полученные в предыдущем разделе, для анализа и оптимизации передачи широковещательных сообщений в локальных сетях IEEE 802.11.

В протоколе IEEE 802.11 регламентирована как однонаправленная передача пакетов, когда пакет предназначается только одной станции-адресату (unicast), так и механизм широковещательной передачи пакета, когда пакет передается всем станциям данной сети (broadcast). Особый интерес представляет широковещательная передача пакетов при распределенном режиме управления, когда нет централизованного управляющего элемента, как в PCF.

В целом и в России, и за рубежом ведутся разработки систем, основанных на широковещательной передаче пакетов в распределенном режиме управления. К примеру, проект «Whizby» [186], где разрабатывается программное обеспечение для карманных компьютеров с радиointерфейсом, которое позволит всем устройствам, находящимся территориально в одном месте, самоорганизовываться в локальную сеть и оповещать о присутствии владельцев этих устройств в данном месте. При этом предполагается, что количество станций в таких сетях может достигать 10 000 в местах массового скопления людей, например, на стадионе.

Отдельным направлением развития беспроводных сетей стали беспроводные распределенные сети датчиков, которые сейчас находят самые разнообразные применения, такие как, например:

- беспроводные датчики слежения за температурой в комнатах дома, передающие информацию на присмник климатической системы, противопожарной охраны, рабочей станции владельца дома [141];
- датчики утечки газопровода на нефтегазовых предприятиях, где измеряемые данные поступают диспетчерам ремонтных служб, причем в таких системах обычно устанавливается сразу несколько приемников сигнала для большей надежности, что обуславливает необходимость широковещательной передачи.

В схеме IEEE 802.11 DCF никаких механизмов подтверждения пакетов с широковещательными сообщениями не предусмотрено. Если необходимо повысить вероятность доставки сообщения, то передающая станция просто отправляет один и тот же кадр несколько раз. Это правило делает необязательным (и даже неэффективным) применение расширенного интервала EIFS, поэтому в данном разделе предполагается, что только интервал DIFS завершает передачу пакета (даже в случае коллизии или искажения кадра помехами). Кроме того, широковещающая станция всегда находится либо на стадии 0 с минимальным конкурентным окном $W = W_0$, либо в состоянии простоя.

6.14.1 Аналитическая модель

Итак, рассмотрим беспроводную сеть, состоящую из N станций и работающую по протоколу IEEE 802.11b, причем каждая станция генерирует широковещатель-

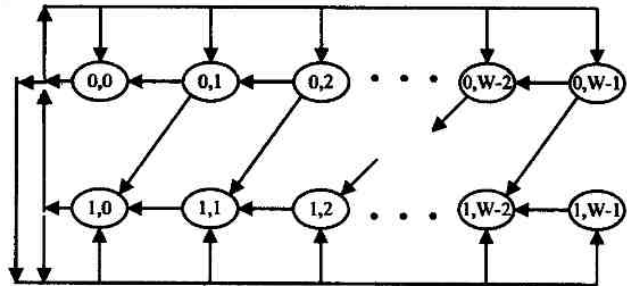
ные пакеты одинаковой длины с интенсивностью λ . Предполагается экспоненциальное распределение времен генерации, и, кроме того, очередь пакетов каждой станции может содержать не более B пакетов, а время распространения сигнала пренебрежимо мало.

Разрабатываемая аналитическая модель направлена на определение оптимальных значений интенсивностей генерации пакетов λ (при заданном количестве станций N), чтобы минимизировать среднее значение $E[T_{\text{ОПОВ}}]$ времени оповещения $T_{\text{ОПОВ}}$, определяемого как интервал между последовательными успешными передачами пакетов одной и той же станцией. Попутно определяются такие важные показатели производительности беспроводной сети, как вероятность коллизий пакетов P_C и вероятность отказа в обслуживании из-за переполнения буфера очереди P_{REL} .

Используя предположения, аналогичные тем, что были сделаны в предыдущем разделе, опишем поведение станции данной сети цепью Маркова на рис. 6.40. Как и ранее, единицей времени рассматриваемого процесса Маркова является виртуальный слот, причем в данном случае возможны лишь следующие три вида слотов:

- 1) «пустой» слот, во время которого ни одна станция не вела передачу;
- 2) асинхронный слот, в котором какая-либо станция вела асинхронную передачу;
- 3) синхронный слот, в котором какая-либо станция пыталась вести синхронную передачу.

Рис. 6.40. Цепь Маркова, изображающая состояния станции



Эта цепь Маркова включает в себя состояния (i, k) , где $i = 0, 1$ — флаг наличия пакетов в очереди, а $k = 0, \dots, W-1$ — значение счетчика отсрочки. В этой цепи состоянию простаивающей станции соответствует состояние $(0, 0)$. Состояния, когда у станции нет пакета для передачи, но при этом она выполняет процедуру отсчета, — это $(0, 1) \dots (0, W-1)$. Наконец, состояния, когда в очереди есть пакеты и станция выполняет процедуру отсрочки, — это $(1, 0), \dots, (1, W-1)$.

Для нахождения стационарных вероятностей $\alpha(i, k)$ определим ненулевые вероятности переходов:

$P\{1, k|1, k+1\} = 1, k \in [0, W-2]$ — уменьшение счетчика отсрочки;

$P\{1, k|1, 0\} = \frac{1-\bar{P}_0}{W}, k \in [0, W-1]$ — синхронная передача, после которой в очереди есть еще хотя бы один пакет;

$P\{0, k|1, 0\} = \frac{\bar{P}_0}{W}, k \in [0, W-1]$ — синхронная передача, после которой в очереди больше нет пакетов;

$P\{1, k|0, k+1\} = P_S$, $k \in [0, W-2]$ — уменьшение счетчика отсрочки, и в пустую очередь приходит пакет;

$P\{0, k|0, k+1\} = 1 - P_S$, $k \in [0, W-2]$ — уменьшение счетчика отсрочки, очередь остается пустой;

$P\{1, k|0, 0\} = \frac{P_S^F + P_S^E P_T}{W}$, $k \in [0, W-1]$ — переход из состояния простоя в состояние отсрочки. Такой переход имеет место, если в момент прихода пакета среда была занята или в момент асинхронной передачи приходит еще один пакет;

$P\{0, k|0, 0\} = \frac{P_S^E(1-P_T)}{W}$, $k \in [1, W-1]$ — переход соответствует асинхронной передаче, после которой в очереди нет больше пакетов и счетчик $b = k > 0$;

$P\{0, 0|0, 0\} = 1 - P_S + \frac{P_S^E(1-P_T)}{W}$ — нет поступивших пакетов или имела место асинхронная передача, за время которой не поступило больше пакетов и счетчик $b = 0$;

$\bar{P}_0 = P_0 e^{-\lambda \text{DIFS}}$, а вероятности P_0 , P_S , P_S^F , P_S^E и P_T были определены в предыдущем разделе.

Определим стационарные вероятности. Исходя из уравнений глобального баланса для состояний $(0, k)$

$$\alpha(0, W-1) = \kappa_0 = \alpha(1, 0) \frac{\bar{P}_0}{W} + \alpha(0, 0) \frac{P_S^E(1-P_T)}{W},$$

$$\alpha(0, W-j) = \kappa_0 + (1-P_S) \alpha(0, W-j+1), \quad j = 2, \dots, W-1,$$

имеем

$$\alpha(0, W-j) = \kappa_0 \sum_{k=0}^{j-1} (1-P_S)^k = \kappa_0 \frac{1 - (1-P_S)^j}{P_S}, \quad j = 1, \dots, W-1.$$

Из уравнений глобального баланса для состояний $(1, k)$

$$\alpha(1, W-1) = \kappa_1 = \alpha(1, 0) \frac{1 - \bar{P}_0}{W} + \alpha(0, 0) \frac{P_S^F + P_S^E P_T}{W},$$

$$\alpha(1, W-j) = \kappa_1 + P_S \cdot \alpha(0, W-j+1) + \alpha(1, W-j+1), \quad j = 2, \dots, W,$$

при $j = 1, \dots, W$ имеем

$$\alpha(1, W-j) = j\kappa_1 + \left(j - \sum_{k=0}^{j-1} (1-P_S)^k \right) \kappa_0 = j(\kappa_1 + \kappa_0) - \kappa_0 \frac{1 - (1-P_S)^j}{P_S}.$$

В частности, для $(1, 0)$ получаем:

$$\alpha(1, 0) = P_S \cdot \alpha(0, 0) + \alpha(1, 0) - \left(\alpha(1, 0) \frac{\bar{P}_0}{W} + \alpha(0, 0) \frac{P_S^E(1-P_T)}{W} \right) \cdot \frac{1 - (1-P_S)^j}{P_S},$$

откуда выражаем $\alpha(1, 0)$ через $\alpha(0, 0)$:

$$\alpha(1, 0) = \frac{\frac{W P_S^2}{1 - (1-P_S)^W} - P_S^E(1-P_T)}{\bar{P}_0} \cdot \alpha(0, 0).$$

Аналогично выражаем остальные стационарные вероятности:

$$\alpha(0, W-j) = P_S \cdot \frac{1 - (1-P_S)^j}{1 - (1-P_S)^W} \cdot \alpha(0, 0), \quad j = 1, \dots, W-1,$$

$$\alpha(1, W-j) = \alpha(0, 0) \left\{ \frac{j}{W} \left(P_S + \frac{\frac{WP_S^2}{1-(1-P_S)^W} - P_S^E (1-P_T)}{\bar{P}_0} \right) - P_S \frac{1-(1-P_S)^j}{1-(1-P_S)^W} \right\},$$

$j = 1, \dots, W$. $\alpha(0, 0)$ определяется из нормировочного уравнения $\sum_{(i,k)} \alpha(i, k) = 1$.

В итоге получаем:

$$\alpha(0, 0) = \left\{ 1 - P_S + \frac{W+1}{2} \cdot \left(P_S + \frac{\frac{WP_S^2}{1-(1-P_S)^W} - P_S^E (1-P_T)}{\bar{P}_0} \right) \right\}^{-1}, \quad (6.92)$$

$$\alpha(1, 0) = \frac{\frac{WP_S^2}{1-(1-P_S)^W} - P_S^E (1-P_T)}{\bar{P}_0} \cdot \alpha(0, 0), \quad (6.93)$$

$$\alpha(0, j) = P_S \cdot \frac{1-(1-P_S)^{W-j}}{1-(1-P_S)^W} \cdot \alpha(0, 0), \quad j = 1, \dots, W-1, \quad (6.94)$$

$$\alpha(1, j) = \frac{W-j}{W} [P_S \cdot \alpha(0, 0) + \alpha(1, 0)] - \alpha(0, j), \quad j = 1, \dots, W-1. \quad (6.95)$$

Пусть τ — вероятность синхронной передачи данной станции за время виртуального слота. Очевидно, $\tau = \alpha(1, 0)$.

Кроме того, аналогично предыдущему разделу находим вероятность того, что синхронная передача станции будет неудачна из-за коллизии, и вероятность асинхронной передачи станции:

$$P_C = 1 - (1 - \tau)^{N-1}, \quad \tau_a = \alpha(0, 0) P_S^E.$$

Определим длительности возможных слотов:

- длительность слота при синхронной передаче $t_S = t_P + \text{DIFS}$, где t_P — время передачи кадра с пакетом;
- длительность слота при асинхронной передаче $t_A = \frac{\sigma}{2} + t_P + \text{DIFS}$;
- длительность пустого слота σ — по определению.

Вероятность прихода пакета за время передачи с учетом DIFS равна $P_T = 1 - \exp(-\lambda t_S)$. Теперь определим вероятности P_S^F и P_S^E . Рассматривая все возможные случаи, аналогично предыдущему разделу получаем:

$$P_S^F = (Q_S + Q_A) P_T, \quad P_S^E = Q_E (1 - e^{-\lambda \sigma}),$$

где

$$Q_S = 1 - (1 - \tau)^{N-1}, \quad Q_A = (N-1) \tau_a (1 - \tau)^{N-2} \text{ и } Q_E = 1 - Q_S - Q_A.$$

Вероятность P_0 опустошения очереди после завершения обслуживания, как и ранее, находится с помощью модели на рис. 6.39 по формуле (6.89). Напомним, что стационарные вероятности π_i состояний этой модели (т.е. вероятности наличия i синхронно обслуживаемых пакетов в очереди) определяются формулами (6.87) и (6.88), а вероятность отказа в приеме пакета из-за переполнения очереди равна $P_{REJ} = \pi_B$.

Оценку вероятности p_a , среднего времени синхронного обслуживания T_S и основного показателя производительности T_{OPOV} проведем в следующем подразделе.

6.14.2 Оценка производительности сети

Как и в предыдущем разделе, для нахождения вероятности p_a и среднего времени синхронного обслуживания разобьем пакеты, обслуживаемые синхронно и поступающие в течение всех возможных виртуальных слотов (i, k) , на следующие четыре категории:

- 1) поступление в течение слотов $(1, k)$, где $k \in [0, W - 1]$;
- 2) поступление в течение слотов $(0, k)$, где $k \in [1, W - 1]$;
- 3) поступление в течение передачи другой станции во время слота $(0, 0)$;
- 4) поступление во время асинхронной передачи данной станции.

Для каждой из этих категорий подсчитаем среднее количество поступающих синхронно обслуживаемых пакетов n_i и n_i^0 .

Тогда время синхронного обслуживания T_S и вероятность p_a определяются по формулам (6.90) и (6.91), компоненты которых, однако, изменяются следующим образом:

$$T_S^* = \frac{W-1}{2} t_{VS} + t_P,$$

где $t_{VS} = Q_E \sigma + Q_S t_S + Q_A t_A$ — средняя длительность виртуального слота, в котором данная станция воздерживается от передачи; кроме того, $\Delta^* = \text{DIFS}$ в (6.90).

Определим значения n_i , n_i^0 и T_S^i для каждой из введенных категорий.

Категория 1:

$$n_1^0 = (1 - e^{-\lambda \cdot \text{DIFS}}) \bar{P}_0 \cdot \alpha(1, 0), \quad n_1 = \lambda t_{VS} \sum_{k=1}^{W-1} \alpha(1, k) + \lambda t_P \cdot \alpha(1, 0),$$

$$T_S^1 = T_S^* + \frac{\text{DIFS}}{2}.$$

Категория 2:

$$n_2^0 = Q^* \sum_{k=1}^{W-1} \alpha(0, k), \quad n_2 = \lambda t_{VS} \sum_{k=1}^{W-1} \alpha(0, k),$$

$$T_S^2 = t_P + \frac{t_{VS} Q^*}{n_2^0} \sum_{k=1}^{W-1} \left(k - \frac{1}{2}\right) \alpha(0, k),$$

где

$$Q^* = Q_E (1 - e^{-\lambda \sigma}) + Q_S (1 - e^{-\lambda t_S}) + Q_A (1 - e^{-\lambda t_A}).$$

Категория 3:

$$n_3^0 = [Q_S (1 - e^{-\lambda t_S}) + Q_A (1 - e^{-\lambda t_A})] \alpha(0, 0), \quad n_3 = \lambda (Q_S t_S + Q_A t_A) \alpha(0, 0),$$

$$T_S^3 = T_S^* + \frac{Q_S t_S + Q_A t_A}{2(1 - Q_E)}.$$

Категория 4:

$$n_4^0 = P_S^E \left(1 - e^{-\lambda(t_P + \text{DIFS})} \right) \alpha(0, 0) = P_T \tau_a, \quad n_4 = \tau_a \lambda (t_P + \text{DIFS}) = \tau_a \lambda \cdot t_S,$$

$$T_S^4 = T_S^* + \frac{t_P + \text{DIFS}}{2} = T_S^* + \frac{t_S}{2}.$$

Окончательно, учитывая асинхронную передачу, получаем среднее время оповещения пакета:

$$E[T_{\text{OPOV}}] = \frac{1}{\lambda [\pi_0 p_a + (1 - \pi_0 p_a) (1 - P_C) (1 - P_{REJ})]}. \quad (6.96)$$

6.14.3 Имитационная модель беспроводной сети

Имитационное моделирование является мощным инженерным методом исследования сложных систем, используемым в тех случаях, когда другие методы оказываются малоэффективными, либо для проверки точности и адекватности аналитических моделей. Имитационная модель представляет собой систему, отображающую структуру и функционирование исходного объекта в виде алгоритма, связывающего входные и выходные переменные, принятые в качестве характеристик исследуемого объекта. Имитационные модели реализуются программно с использованием различных языков. Одним из наиболее распространенных языков, специально предназначенных для построения имитационных моделей, является GPSS: General Purpose Simulation System (моделирующая система общего назначения) [157, 195]. Этот язык разработан фирмой IBM и входит в число наиболее распространенных языков программирования.

Система GPSS ориентирована на класс объектов, процесс функционирования которых можно представить в виде множества состояний и правил перехода из одного состояния в другое, определяемых в дискретной пространственно-временной области. Примерами таких объектов являются вычислительные системы, сети ЭВМ, системы передачи сообщений, транспортные объекты, склады, магазины, предприятия и т. п. В качестве формальных моделей таких объектов используют системы массового обслуживания, автоматы, стохастические сети, сети Петри и макросети, агрегаты и т. п.

Разработчик конструирует модель из блоков, прибегая, как правило, к наглядной форме ее отображения в виде блок-схемы. Для удобства графического представления модели каждый блок GPSS имеет принятое стандартное обозначение. Построенная схема является одновременно программой на языке GPSS. Для ее ввода в ЭВМ необходимо последовательность блоков представить в виде списка операций, добавив к названиям блоков требуемые операнды.

Функционирование объекта отображается в модели в виде перемещения транзактов от блока GENERATE в блок TERMINATE через промежуточные блоки. Транзакты, или сообщения, являются абстрактными подвижными элементами, которые могут моделировать различные объекты реального мира: сообщения, программы, транспортные средства, людей и т. п. Перемещаясь между блоками модели, транзакты вызывают (и испытывают) различные действия. Возможны их задержки в некоторых точках модели, изменения маршрутов и направлений движения, расщепление транзактов на несколько копий и т. п. С каждым тран-

заком связан упорядоченный набор данных. Он включает номер транзакта; номер блока, в котором в данный момент находится транзакт; номер следующего блока; время перехода в следующий блок; приоритет, характеризующий очередность обработки транзактов в определенных случаях; а также набор параметров, с помощью которых каждому транзакту можно присвоить числовые значения, выражающие желаемые свойства или характеристики моделируемых объектов: вес, скорость, объем, цвет, время обработки и т. п.

Как и в аналитической модели, предположим, что все станции, находящиеся в состоянии задержки, изменяют свои счетчики задержки сразу после интервала DIFS, завершающего передачу пакета (в действительности станции с ненулевым значением счетчиков задержки изменяют их только спустя σ с момента окончания интервала DIFS). Таким образом, в начале каждого виртуального слота любая станция, имеющая пакеты в своей очереди, может предпринять синхронную попытку передачи.

Итак, рассмотрим беспроводную локальную сеть, состоящую из $N = 50$ станций. Каждая станция представляется массивом данных, в котором хранятся параметры, полностью описывающие поведение станции.

Таблица 6.7. Параметры имитационной модели

обозначение	значения	описание
NUM	1, ..., 50	Номер станции
BOFF	0, ..., $W - 1$	Значение счетчика отсрочки
Q	0, ..., B	Число пакетов в очереди станции
TLAST		Время окончания последней успешной передачи пакета
TIN		Время поступления следующего пакета в очередь

Модельное время обозначим через $C1$. Время оповещения ($TOPOV$) определим как разность времени на момент окончания успешной передачи текущего пакета станции и времени $TLAST$: $TOPOV = C1 - TLAST$. Значения $TOPOV$ будем сохранять в массив результатов функцией GPSS TABULATE.

Процесс моделирования заключается в следующем: на границе окончания предыдущего виртуального слота, исходя из заданных в массивах параметров, определяются события, происходящие в следующем виртуальном слоте, и в соответствии с этим изменяются параметры всех станций. Процессы, происходящие в сети в течение одного виртуального слота, описываются с большой точностью следующим алгоритмом.

Могут иметь место следующие случаи:

1. *Попытка синхронной передачи.*

Просматриваем счетчики отсрочки станций с непустой очередью $Q > 0$. Если $BOFF = 0$ хотя бы для одной станции, то длительность виртуального слота $TVS = t_P + DIFS$. Если $BOFF = 0$ только для одной станции, то имеет место успешная синхронная передача.

Определяем время оповещения: $TOPOV = C1 - TLAST + TVS - DIFS$,
 $TABULATE TOPOV$.

Пересчитываем $TLAST = C1 + TVS - DIFS$.

Если $BOFF = 0$ сразу для нескольких станций, то имеет место коллизия.
 Для передававших(ей) станций:

- разыгрываем значения счетчиков отсрочки $BOFF$,
- уменьшаем очереди передававших станций $Q = Q - 1$.

Переходим к пункту 4.

Иначе, если $BOFF > 0$ для всех станций с непустой очередью, то переходим к следующему случаю.

2. *Асинхронная передача.*

Рассматриваем станции с пустой очередью $Q=0$ и нулевым счетчиком отсрочки $BOFF = 0$.

Если $TIN - C1 \leq \sigma$ для одной из этих станций, то в виртуальном слоте происходит асинхронная передача.

Определяем длительность виртуального слота:

$$TVS = (C1 - TIME) + t_p + DIFS.$$

Время оповещения

$$TOPOV = C1 - TLAST + TVS - DIFS, \quad TABULATE TOPOV.$$

Пересчитываем время последней успешной передачи

$$TLAST = TIME + TVS - DIFS.$$

Переопределяем время TIN для передававшей станции, разыгрываем время отсрочки $BOFF$ для передававшей станции и переходим к пункту 4.

Если $TIN - C1 > \sigma$ для всех станций, тогда рассматриваем последний случай.

3. *Пустой слот.*

Определяем длительность виртуального слота: $TVS = \sigma$. Переходим к следующему пункту.

4. *Пересчет состояний очередей за время виртуального слота.*

Время, оставшееся до прихода следующего пакета в очередь $TIN - C1$, сравнивается со временем виртуального слота TVS .

Если $TIN - C1 \leq TVS$, то а) если $Q = 0$, то $Q = Q + 1$ и, если при этом $BOFF = 0$, то разыгрываем значение времени отсрочки; б) если $0 < Q < B$, то $Q = Q + 1$.

Разыгрываем момент прихода следующего пакета TIN в соответствии с экспоненциальным законом распределения и снова переходим к пункту сравнения $TIN - C1 \leq TVS$, чтобы учесть возможность прихода нескольких пакетов за время виртуального слота.

Как только $TIN - C1 > TVS$, пересчет состояния очереди для данной станции окончен.

5. Уменьшаем счетчики станций, которые не передавали в данном виртуальном слоте и у которых параметр $BOFF > 0$: $BOFF = BOFF - 1$.

6. Текущее модельное время $C1$ переводим на момент окончания виртуального слота: $C1 = C1 + TVS$.

Переходим к пункту 1.

Численные результаты. Имитационное моделирование использовалось для обоснования работоспособности аналитической модели. На рис. 6.41 приведена зависимость среднего времени оповещения $E[T_{OPOV}]$ от среднего времени генерации, полученная независимо обеими моделями, при количестве станций в локальной беспроводной сети $N = 50$ и размере пакета данных $L = 1000$ байт, что соответствует времени передачи кадра $t_P = 850$ мкс. Результаты имитационного моделирования полностью подтверждают работоспособность аналитической модели.

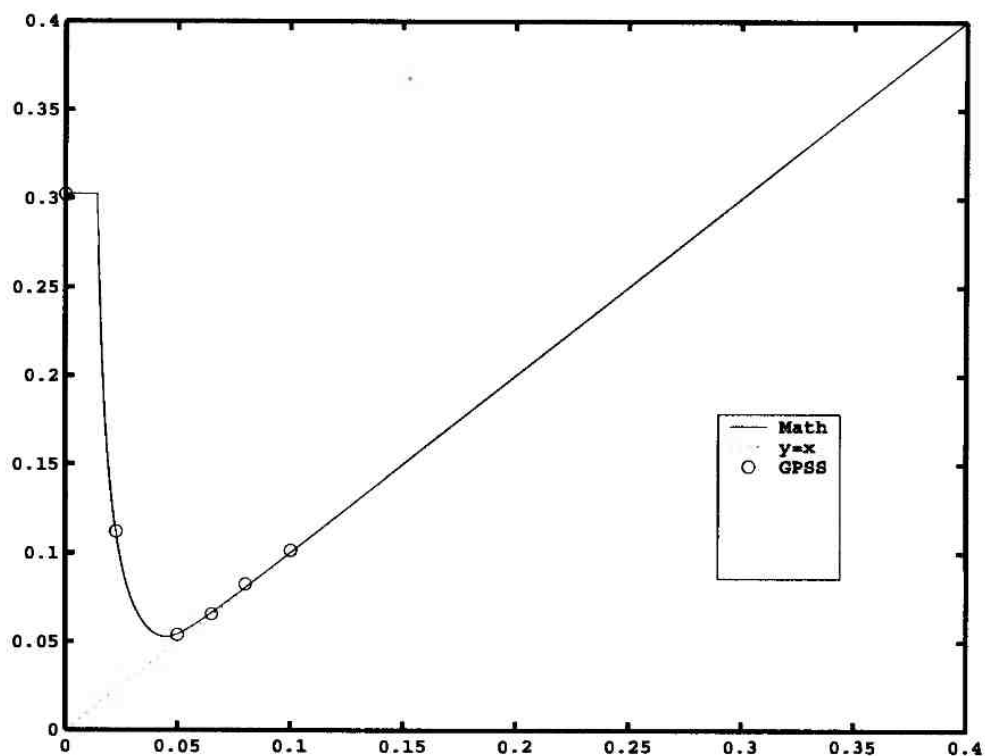


Рис. 6.41. Зависимость среднего времени оповещения $E[T_{OPOV}]$ (с) от среднего времени генерации λ^{-1} (с), полученные аналитически (Math) и имитационно (GPSS)

На графике явно выражено минимальное значение среднего времени оповещения. Его возрастание при увеличении среднего времени генерации обуславливается тем, что все пакеты передаются асинхронно, а значит, успешно, и в этом случае время оповещения в точности равно времени генерации. Возрастание $E[T_{OPOV}]$ при уменьшении времени генерации обуславливается процессами коллизий, так как большая часть пакетов передается синхронно, и отказами в обслуживании вследствие переполнения очереди. Случаю насыщения, когда в очередях станций всегда есть пакеты, соответствует горизонтальный участок графика.

6.15 Оценка пропускной способности беспроводной сети типа «hot spot» и ее оптимизация

6.15.1 Сети типа «hot spot»

В предыдущих разделах данной главы мы в основном рассматривали так называемые самоорганизующиеся (ad hoc) сети. Однако чаще всего беспроводные локальные сети используются именно для предоставления услуг мобильного беспроводного доступа к информационным ресурсам (такие сети часто называют общественными точками доступа, сетями типа «hot spot» или зонами «Wi-Fi»). Во всем мире беспроводной доступ приобретает все большую популярность: беспроводные сети развертываются в аэропортах, гостиницах, торговых комплексах — везде, где может потребоваться быстрый мобильный доступ к информации. С развитием рынка телекоммуникаций данная услуга становится все более доступной и востребованной в силу неизменно растущей потребности общества в информационном обмене.

При проектировании беспроводных сетей общественного доступа приходится учитывать множество факторов, необходимых для успешного и эффективного построения сети, предоставления различных услуг доступа к данным и гарантирования соответствующего качества обслуживания. Для конечного пользователя основную роль будут играть реальная скорость передачи данных, доступные сервисы передачи информации — передача данных, голоса (Voice over IP — VoIP), видеоконференции, а также качество и надежность связи. Поэтому для беспроводных сетей в силу их специфики особо остро стоит вопрос обеспечения заданного уровня качества сервиса (Quality of Service — QoS). Параметры пропускной способности являются одной из основных характеристик QoS. Реальная емкость канала передачи данных, предоставляемого пользователю, зависит от используемых сетевых протоколов, внешних условий, в которых развертывается сеть (наличие физических препятствий и источников шума, уровни затухания сигнала), количества активных пользователей в сети, типа информации (мультимедийный трафик, данные) и характера ее передачи. Именно поэтому перед непосредственным развертыванием беспроводной сети должно быть проведено теоретическое исследование ее работы.

В связи с этим целью данного раздела является оценка пропускной способности беспроводной сети типа «hot spot» и нахождение оптимальных параметров сети для обеспечения равных условий и качества беспроводного доступа для всех станций «hot spot»-сети.

При организации сети публичного доступа, как правило, используется протокол IEEE 802.11b. Будучи утвержденным в качестве стандарта еще в 1999 году, он за несколько лет приобрел наибольшее распространение. Беспроводные сети, как и их проводные аналоги, являются частью глобальной сети Internet и, следовательно, должны иметь аналогичные возможности по обеспечению качества сервиса. Однако в стандарте 802.11b отсутствуют какие-либо средства поддержки QoS. В итоге была начата работа над новым протоколом IEEE 802.11e [176],

который при сохранении совместимости с уже принятыми стандартами обеспечил бы расширение их функциональности за счет поддержки мультимедийных потоков данных и гарантированного качества услуг. Первая версия спецификации 802.11e была представлена в конце 2001 года, а утверждение этого протокола в качестве самостоятельного стандарта планировалось на лето 2004 года. Новый протокол обладает гораздо большей гибкостью в настройке его параметров по сравнению со своими предшественниками. Эта гибкость позволяет управлять качеством обслуживания на уровне доступа к среде передачи данных силами самого протокола, без использования дополнительных технологий. Но из-за новизны технологии и недостаточного ее теоретического исследования производители сетевого оборудования не спешат реализовывать аппаратно возможности по настройке параметров протокола 802.11e.

В данном разделе показывается эффективность управления пропускной способностью канала передачи данных путем изменения параметров используемого протокола и обосновывается целесообразность аппаратной реализации такой функциональности.

Как уже говорилось, чаще всего «hot spot»-сети развертываются в общественных местах, где необходим быстрый доступ к сети Internet.

Основными компонентами такой сети являются терминальные станции (ТС) и базовая станция (БС). ТС нестационарны, т.е. могут в процессе функционирования сети подключаться к ней или отключаться от нее, перемещаться в ее пределах и покидать их. ТС представляют собой различные портативные и мобильные устройства, имеющие сетевой адаптер с поддержкой протоколов IEEE 802.11: ноутбуки, карманные компьютеры, мобильные телефоны. ТС «общаются» только с БС, играющей в этом случае роль шлюза между беспроводной и внешней сетями.

Радиус зоны покрытия типичной БС составляет несколько сотен метров при прямой видимости (при большем расстоянии до БС начинает сказываться затухание сигнала, и скорость передачи данных резко падает). Этим значением и ограничены размеры «hot spot»-сети с одной БС. Можно развертывать беспроводные сети и большей протяженности, используя несколько БС, организуя перекрывающиеся зоны обслуживания. Мобильная станция, оказавшаяся в области действия нескольких БС, регистрируется в одной из зон обслуживания и продолжает затем работу только с одной из БС.

Число пользователей (соответственно, и количество ТС) беспроводной сети обычно составляет несколько десятков человек. При этом, конечно, нужно учитывать среднее число активных абонентов, определяющее в конечном счете пропускную способность всей сети и реальную доступную пользователю скорость передачи данных.

Помимо предоставления доступа к всемирной сети беспроводный «hot spot»-доступ может потребоваться при организации различных телеконференций, услуг передачи голоса и изображения.

В большинстве беспроводных сетей для передачи данных используется только один частотный канал. Но в последнее время на рынке оборудования беспроводного доступа начали появляться устройства, поддерживающие два независимых частотных канала — один для передачи от БС к ТС, а другой для передачи

от ТС к БС. В этом случае БС будет занимать всю возможную емкость выделенного ей канала. Соответственно, ТС будут использовать для своей передачи другой канал, совместно эксплуатируя его ресурсы. В этом случае говорят, что сеть работает в *полнодуплексном* режиме («full-duplex»).

6.15.2 Математическая модель сети

Рассмотрим беспроводную сеть, включающую N статистически однородных ТС и БС. Учитываем только *активные*, работающие в *насыщении* станции — т.е. станции, в очереди которых всегда находится, по крайней мере, один пакет. Как и ранее, под статистической однородностью станций понимается одинаковость распределений $\{d_\ell, \ell = \ell_{\min}, \dots, \ell_{\max}\}$ длин пакетов, поступающих из очереди. Сеть имеет типичную для «hot spot»-систем физическую топологию «сота»: в центре соты находится БС, вокруг которой располагаются ТС. ТС общаются только с БС, но при этом «слышат» друг друга (т.е., отсутствуют скрытые станции). Полагаем также, что весь трафик от ТС к БС направлен во внешнюю сеть, а от БС к ТС — из внешней сети, т.е. между абонентами беспроводной сети отсутствуют информационные потоки. Кроме того, предполагается равномерность распределения трафика от БС ко всем ТС, т.е. текущий пакет БС предназначен любой ТС i с вероятностью $1/N$. Из внешних факторов рассматриваются помехи и *затухание* сигнала. Влияние затухания учитывается следующим образом: поскольку размеры территории покрытия «hot spot»-сети составляют несколько десятков метров, то предположение об отсутствии скрытых станций справедливо, но остается возможным «перекрывание» одного сигнала другим. Поясним это.

Пусть две станции находятся на расстояниях $R_{[1]}$ и $R_{[2]}$ ($R_{[2]} > R_{[1]}$) от БС (см. рис. 6.42). Тогда при одновременной передаче (в системе без затухания возникла бы коллизия) БС успешно воспримет сигнал от станции № 2 как шум на фоне сигнала от станции № 1, если отношение этих сигналов больше некоторой величины A [дБ], т.е.

$$10 \cdot \lg \left(\frac{R_{[2]}^\nu}{R_{[1]}^\nu} \right) > A, \quad (6.97)$$

где $\nu = 2$ для вакуума или коротких дистанций (комнатных сетей) и $\nu \approx 3,3$ для городских условий [133].

Таким образом, станция ТС $_{[i]}$ с радиус-вектором $R_{[i]}$ будет *конкурировать* только с теми станциями, которые расположены не дальше чем на расстоянии $R'_{[i]} = R_{[i]} \cdot 10^{A/10\nu}$ от БС. Именно эти станции определяют вероятность коллизии и вероятность передачи для ТС $_{[i]}$.

БС является стационарным устройством, поэтому ее выходная мощность обычно в несколько раз больше мощностей передатчиков мобильных пользователей.

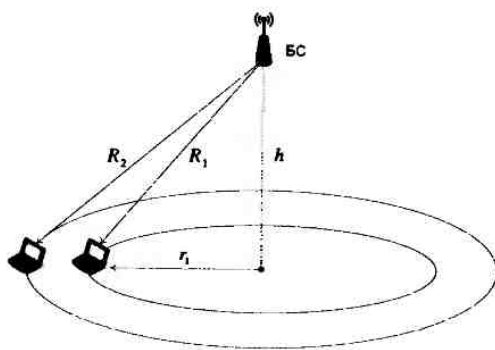


Рис. 6.42. Схема сети «hot spot»

Следовательно, можно считать, что БС «перекрывает» сигналы всех других станций. Тогда каждая ТС сможет распознать передачу БС.

Как и в предыдущих разделах, разделим все время работы сети на виртуальные временные слоты различной длины. В начале каждого такого слота станция уменьшает значение счетчика отсрочки и, если этот счетчик обнуляется, начинает передачу.

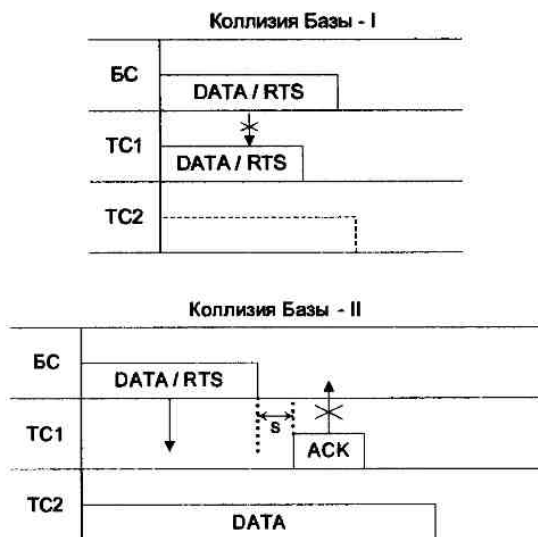
Виртуальный слот может представлять собой:

- «пустой» слот, в который ни одна из станций не ведет передачу;
- «успешный» слот, когда происходит успешная передача пакета данных (и подтверждение приема) от БС к ТС или от ТС к БС;
- «коллизийный» слот, когда одновременно передают две станции и более, и передаваемые кадры искажаются из-за взаимного наложения друг на друга.

Пусть в некоторый временной слот БС начала передачу, тогда для нее этот слот будет «коллизийным» в следующих случаях (см. рис. 6.43):

- станция назначения ($ТС_{[1]}$) также ведет передачу;
- одновременно с БС передает одна из станций ($ТС_{[2]}$), с которыми конкурирует $ТС_{[1]}$, причем $t_f^{ТС_{[2]}} > t_f^{БС} + SIFS$ (t_f — время передачи кадра) — в этом случае искажается кадр АСК, посылаемый $ТС_{[1]}$.

Рис. 6.43. Типы коллизий БС



Во всех остальных случаях данный виртуальный слот будет «успешным» для БС.

Аналогично, для некоторой ТС виртуальный слот будет «коллизийным», если одновременно передают либо БС, либо одна из конкурентных к данной ТС станций, и «успешным» в противном случае.

Будем также предполагать, что вероятность начала передачи данной станцией в данном слоте не зависит от предыстории работы сети. Тогда вероятности того, что произвольно выбранный виртуальный слот будет «пустым» (p_e),

«успешным» (p_s) или «коллизийным» (p_c), определяются выражениями:

$$p_e = \prod_{i=0}^N (1 - \tau_{[i]}), \quad p_s = \sum_{i=0}^N p_{s[i]}, \quad p_c = 1 - p_e - p_s. \quad (6.98)$$

Здесь $p_{s[i]}$ — вероятность «успешного» слота для некоторой i -й станции ($n_{[i]}$ — множество терминальных станций, с которыми она конкурирует, причем $n_{[i]} = \emptyset$), а $\tau_{[i]}$ — вероятность начала ее передачи в произвольно выбранном виртуальном слоте. Все параметры с «нулевыми» индексами относятся к БС. Далее,

$$p_{s[0]} = \tau_{[0]} \sum_{i=1}^N \frac{1 - \tau_{[i]}}{N} \sum_{\ell_0=\ell_{\min}}^{\ell_{\max}} \hat{d}_{\ell_0[0]} \prod_{k \in n_{[i]}} \left[1 - \tau_{[k]} \cdot \sum_{\ell_k > \ell} \hat{d}_{\ell_k[k]} \right],$$

$$p_{s[i]} = \tau_{[i]} (1 - \tau_{[0]}) \prod_{k \in n_{[i]}} (1 - \tau_{[k]}), \quad i = 1, \dots, N, \quad (6.99)$$

где $\hat{d}_{\ell[k]}$ — вероятность передачи пакета длиной ℓ станцией k . Важно, что распределение $\{\hat{d}_{\ell}, \ell = \ell_{\min}, \dots, \ell_{\max}\}$ отличается от $\{d_{\ell}, \ell = \ell_{\min}, \dots, \ell_{\max}\}$. d_{ℓ} — это распределение длин пакетов, поступающих на MAC-уровень от протоколов более высоких уровней, в то время как $\hat{d}_{\ell[k]}$ определяет вероятность того, что в рассматриваемую попытку передачи k -й станции происходит передачи пакета длиной ℓ . Отличие распределений d_{ℓ} и $\hat{d}_{\ell}$ связано с наличием в сети внешнего шума, ведь чем длиннее пакет, тем больше вероятность его искажения и тем больше число попыток его передачи.

Таким образом, пропускная способность S всей сети находится по формуле:

$$S = \frac{p_s \bar{U}}{p_e \sigma + p_s \bar{T}_s + p_c \bar{T}_c}, \quad (6.100)$$

где $\bar{T}_s$ и $\bar{T}_c$ — средние длительности «успешного» и «коллизийного» слотов, а $\bar{U}$ — среднее число байт информации, переданной в течение «успешного» слота.

Можно также выделить пропускную способность канала «снизу вверх» (от ТС к БС — S_{up}) и «сверху вниз» (от БС к ТС — S_{down}):

$$S = S_{up} + S_{down}; \quad S_{up} = \sum_{i=1}^N S_{up}^{[i]}, \quad S_{down} = \frac{p_{s[0]} \bar{U}_{[0]}}{p_e \sigma + p_s \bar{T}_s + p_c \bar{T}_c}, \quad (6.101)$$

где

$$S_{up}^{[i]} = \frac{p_{s[i]} \bar{U}_{[i]}}{p_e \sigma + p_s \bar{T}_s + p_c \bar{T}_c} \quad (6.102)$$

— емкость канала, приходящаяся на ТС i , а $\bar{U}_{[i]}$ аналогично $\bar{U}$, но для этой конкретной ТС.

Считая, что вероятности коллизии трех станций и более пренебрежимо малы по сравнению с вероятностью коллизии двух станций, получим:

$$\bar{T}_c = \frac{1}{\sum_{i,j \in n_{[i]}} p_c^{ij}} \cdot \sum_{i,j \in n_{[i]}} p_c^{ij} \bar{T}_c^{ij}, \quad (6.103)$$

где

$$p_c^{ij} = p_{s[i]} \cdot \frac{\tau_{[j]}}{1 - \tau_{[j]}}$$

— вероятность одновременной передачи только лишь станций i и j ;

$$\begin{aligned} \overline{T_c^{ij}} = & \sum_{\ell=\ell_{\min}}^L t(\ell) \left[\left(\sum_{k=\ell_{\min}}^{\ell} \hat{d}_{k[i]} + \sum_{k=L+1}^{\ell_{\max}} \hat{d}_{k[i]} \right) \left(\sum_{k=\ell_{\min}}^{\ell} \hat{d}_{k[j]} + \sum_{k=L+1}^{\ell_{\max}} \hat{d}_{k[j]} \right) - \right. \\ & - \left(\sum_{k=\ell_{\min}}^{\ell-1} \hat{d}_{k[i]} + \sum_{k=L+1}^{\ell_{\max}} \hat{d}_{k[i]} \right) \left(\sum_{k=\ell_{\min}}^{\ell-1} \hat{d}_{k[j]} + \sum_{k=L+1}^{\ell_{\max}} \hat{d}_{k[j]} \right) \Big] + \\ & + t_{\text{RTS}} \left(\sum_{k=L+1}^{\ell_{\max}} \hat{d}_{k[i]} \right) \left(\sum_{k=L+1}^{\ell_{\max}} \hat{d}_{k[j]} \right) + \text{EIFS}, \end{aligned}$$

— средняя длительность коллизии станций i и j , а L — RTS-порог. В квадратных скобках последней формулы стоит дифференциал функции распределения общей длительности передачи станций (с момента начала до момента окончания передачи последней станции) при условии, что хотя бы одна из станций использует в своей попытке передачи базовый метод доступа. Здесь учтен тот факт, что функция распределения максимума случайных величин есть произведение функций распределения этих величин. Кроме того, в выражении для $\overline{T_c^{ij}}$ $t(\ell) = t_H + \ell/V$ — время передачи кадра с данными, t_H — время передачи заголовка кадра и ℓ — длина пакета данных.

Пусть в начале «успешного» слота одна из станций начала передачу пакета длиной ℓ , тогда ее попытка не удастся со следующей вероятностью:

$$\text{для ТС:} \quad \pi_h(\ell) = \begin{cases} \xi_d(\ell), & \ell \leq L, \\ 1 - (1 - \xi_r)(1 - \xi_d(\ell)), & \ell > L; \end{cases} \quad (6.104)$$

$$\text{для БС:} \quad \pi_h(\ell) = \begin{cases} \xi_a, & \ell \leq L, \\ 1 - (1 - \xi_a)^2, & \ell > L; \end{cases} \quad (6.105)$$

где вероятности искажения ξ определены в подразделе 6.10.1.

Тогда средняя длительность «успешного» слота при передаче i -й станцией пакета длиной ℓ равна:

$$\overline{T_s(\ell)} = t(\ell) + [1 - \xi_d(\ell)] (\text{SIFS} + t_{\text{ACK}}) + [1 - \pi_h(\ell)] \text{DIFS} + \pi_h(\ell) \text{EIFS}, \quad \ell \leq L, \quad (6.106)$$

$$\begin{aligned} \overline{T_s(\ell)} = & t_{\text{RTS}} + (1 - \xi_r) (t_{\text{CTS}} + \text{SIFS}) + \\ & + (1 - \xi_r) \{t(\ell) + \text{SIFS} + [1 - \xi_d(\ell)] (t_{\text{ACK}} + \text{SIFS})\} + \\ & + [1 - \pi_h(\ell)] \text{DIFS} + \pi_h(\ell) \text{EIFS}, \quad \ell > L. \end{aligned} \quad (6.107)$$

Аналогично для БС:

$$\overline{T_s(\ell)} = t(\ell) + \delta + (\text{SIFS} + t_{\text{ACK}} + \delta) + [1 - \pi_h(\ell)] \text{DIFS} + \pi_h(\ell) \text{EIFS}, \quad \ell \leq L; \quad (6.108)$$

$$\begin{aligned} \overline{T_s(\ell)} = & t_{\text{RTS}} + (t_{\text{CTS}} + \text{SIFS}) + \\ & + (1 - \xi_a) \{t(\ell) + \text{SIFS} + (t_{\text{ACK}} + \text{SIFS})\} + \\ & + [1 - \pi_h(\ell)] \text{DIFS} + \pi_h(\ell) \text{EIFS}, \quad \ell > L. \end{aligned} \quad (6.109)$$

В итоге, получаем:

$$\overline{T_s} = \frac{1}{p_s} \sum_{i=0}^N p_{s[i]} \overline{T_{s[i]}}, \quad \overline{T_{s[i]}} = \sum_{\ell=\ell_{\min}}^{\ell_{\max}} \overline{T_s(\ell)} \hat{d}_{\ell[i]}; \quad (6.110)$$

$$\overline{U} = \frac{1}{p_s} \sum_{i=0}^N p_{s[i]} \overline{U_{[i]}}, \quad \overline{U_{[i]}} = \sum_{\ell=\ell_{\min}}^{\ell_{\max}} (1 - \pi_h(\ell)) \ell \hat{d}_{\ell[i]}. \quad (6.111)$$

Таким образом, получены все компоненты формулы (6.100), и пропускная способность может быть найдена, если известны вероятности передачи $\tau_{[i]}$ и распределения $\{\hat{d}_{\ell[i]}\}$ длин пакетов, передаваемых станциями. В свою очередь, $\tau_{[i]}$ и $\{\hat{d}_{\ell[i]}\}$ определяются по формулам раздела 6.10, в которых только нужно использовать вероятность π_c неудачи из-за коллизии в виде

$$\pi_c = 1 - (1 - \tau_{[0]}) \prod_{k \in n_{[i]}} (1 - \tau_{[k]}) \quad (6.112)$$

для ТС_[i] (т. е. при определении $\tau_{[i]}$) и

$$\pi_c = 1 - p_{s[0]} / \tau_{[0]} \quad (6.113)$$

для БС (т. е. при определении $\tau_{[0]}$).

6.15.3 Оптимизация пропускной способности

Ввиду присутствия в системе затухания сигнала каждая станция будет обладать своими собственными вероятностью передачи и пропускной способностью. Причем чем дальше находится станция от БС, тем меньше будет доступная ей реальная емкость канала передачи, что, естественно, несправедливо. Попробуем, подобрав соответствующие параметры станций (минимальное значение конкурентного окна $CW_{\min}$), уменьшить $\tau_{[i]}$ для станций, находящихся вблизи БС. Тогда пропускная способность для станций, находящихся около БС, уменьшится, а для тех, что расположены вдали от БС, возрастет. В итоге получим беспроводную сеть с равным качеством обслуживания, т. е. с фиксированной емкостью канала доступа в Интернет для любой станции, независимо от ее физического расположения относительно БС.

Далее излагается метод оценки пропускной способности сети при равном качестве обслуживания для всех ТС. Как уже говорилось ранее, для каждой ТС существует определенное число «конкурентных» станций. Назовем *слоем* некоторое множество станций, обладающих равным количеством ТС, конкурирующих с ними. Все станции одного слоя $m = 1, \dots, n$ будут обладать равными вероятностями передачи τ_m и, соответственно, равными пропускными способностями.

Рассмотрим станцию, принадлежащую некоторому m -му слою; вероятность «успешного» слота для нее равна:

$$p_s(m) = \frac{\tau_m}{1 - \tau_m} \left[\prod_{j=1}^n (1 - \tau_j)^{\varphi(\tau_m, j)} \right] (1 - \tau_{[0]}),$$

где

$$\varphi(m, j) = \begin{cases} k_j, & N_m \geq \sum_{h=1}^j k_h; \\ N_m - \sum_{h=1}^{j-1} k_h, & \sum_{h=1}^{j-1} k_h < N_m < \sum_{h=1}^j k_h; \\ 0, & N_m \leq \sum_{h=1}^{j-1} k_h. \end{cases} \quad (6.114)$$

Здесь $N_m - 1$ — число конкурентных станций для ТС m -го слоя, а k_j — количество станций, принадлежащих слою j .

Для станции, принадлежащей $(m+1)$ -му слою, аналогичная вероятность успешной передачи равна $p_s(m+1)$. Тогда:

$$\frac{p_s(m+1)}{p_s(m)} = \tau_{m+1} \cdot \frac{1 - \tau_m}{\tau_m} \cdot \left[\frac{\prod_{j=m+2}^n (1 - \tau_j)^{\varphi(m+1, j)}}{\prod_{j=m+1}^n (1 - \tau_j)^{\varphi(m, j)}} \right] \cdot (1 - \tau_{m+1})^{k_m - 1}. \quad (6.115)$$

Поскольку величина $\bar{U}_i$ практически постоянна для всех станций (согласно полученным результатам различие может достигать 3% при количестве станций ~ 100), то для равенства пропускных способностей необходимо, чтобы

$$\frac{p_s(m+1)}{p_s(m)} \approx 1. \quad (6.116)$$

Таким образом, обозначив

$$\left[\frac{\prod_{j=m+2}^n (1 - \tau_j)^{\varphi(m+1, j)}}{\prod_{j=m+1}^n (1 - \tau_j)^{\varphi(m, j)}} \right] \cdot (1 - \tau_{m+1})^{k_m - 1} = \mathfrak{S}_m, \quad \mathfrak{S}_m = \mathfrak{S}_m(\tau_{m+1}, \dots, \tau_n),$$

получаем рекуррентную формулу для определения вероятности передачи станций из m -го слоя:

$$\tau_m = \frac{\tau_{m+1} \cdot \mathfrak{S}_m}{1 + \tau_{m+1} \cdot \mathfrak{S}_m}. \quad (6.117)$$

Работа сети в режиме «full-duplex». Пусть теперь беспроводная сеть построена на оборудовании, работающем в «full-duplex»-режиме. Оценку пропускной способности частотных каналов в этом случае нужно производить независимо друг от друга.

Таким образом, для канала передачи терминальных станций исключается из рассмотрения БС, полагая, например, во всех приведенных выше формулах $\tau_{[0]} = 0$.

Далее, при оценке нагрузки на канал передачи БС учитывается передача только самой БС. Вероятность передачи БС можно также найти по формулам раздела 6.10 при условии, что вероятность коллизии при передаче БС равна нулю, т.е. $\pi_c = 0$. Вероятности виртуальных слотов будут определяться только лишь вероятностью передачи БС:

$$p_e = 1 - \tau_{[0]}, \quad p_s = \tau_{[0]}, \quad p_c = 0,$$

и пропускная способность базового канала может быть вычислена по формуле

$$S_{down} = \frac{p_s \overline{U}_{[0]}}{p_e \sigma + p_s \overline{T}_s} = \frac{\tau_{[0]} \overline{U}_{[0]}}{(1 - \tau_{[0]}) \sigma + \tau_{[0]} \overline{T}_s}, \quad (6.118)$$

где среднее время $\overline{T}_s$ успешного слота и среднее число байт успешно переданной информации определяются формулами (6.105), (6.108)–(6.111).

Процедуры оценки пропускной способности. Для оценки пропускной способности сети применяем следующий итеративный алгоритм.

1. Задаем параметры как всей системы, так и отдельных станций (например, размер минимального конкурентного окна $CW_{min} \equiv W_0$); определяем начальные значения τ_i (например, $\tau_i = 1/CW_{min}$, $i = 1, N$).

2. Для каждой станции и БС, всевозможных длин пакетов ℓ и количества попыток i вычисляем вероятности $\psi_\ell(i)$, средние $\overline{f}_\ell$ и $\overline{w}_\ell$.

3. Находим новые значения вероятности передачи τ_i и сравниваем их с предыдущими. Если разница между ними больше некоторого значения — точности оценки τ_i , то возвращаемся к шагу 2 с τ_i , равными полусумме старого и нового значений.

Расчет пропускной способности при оптимальных параметрах, реализующих равные пропускные способности для каждой станции, проводим следующим образом.

1. Вычисляем τ для наиболее удаленной станции, взяв в качестве CW_{min} минимальное значение ($CW_{min} = 32$) и используя для оценки вероятностей передачи других станций выражение (6.116). Получаем в итоге значения τ для всех станций.

2. Перебирая для каждой из оставшихся станций минимальное значение CW_{min} , вычисляем соответствующее ей τ и добиваемся его максимального соответствия тому, что было получено на шаге 1.

3. Проводим перерасчет пропускных способностей станций уже для новых значений вероятностей передачи.

6.15.4 Численные оценки

Численные оценки проводились для беспроводной сети, включающей N статистически однородных ТС и БС, полагая, что данная сеть работает под управлением протокола IEEE 802.11e с использованием схемы DCF. Значения параметров протокола, используемые для получения численных результатов аналитической модели, представлены в табл. 6.4. Распределение пакетов $\{d_\ell\}$ принималось равномерным в интервале $[1, \dots, 1999]$.

При расчетах полагалось, что все терминальные станции находятся в одной и той же горизонтальной плоскости и располагаются на концентрических окружностях радиусом 1, 2, 3 м и т. д. Число ТС на окружности выбиралось так, чтобы станции располагались равномерно, с одной и той же плотностью λ по всей обслуживаемой территории.

Все численные результаты были получены для сети со следующими параметрами: расстояние от БС до плоскости расположения станций $h = 2$ м; плотность

расположения станций $\lambda = 0,2 \text{ м}^{-2}$ (соответствующее расположение станций показано в табл. 6.8); коэффициент ослабления радиосигнала при распространении $\nu = 3,3$; порог обнаружения сигнала $A = 10 \text{ дБ}$.

Таблица 6.8. Число станций n_r на окружностях радиуса r

r	1	2	3	4	5	6	7	8	9	10
n_r	0	1	1	1	2	2	2	3	4	4
r	11	12	13	14	15	16	17	18	19	20
n_r	4	4	5	6	6	6	6	7	8	8

Используя приведенные выше данные и соотношение (6.97), для каждой станции был вычислен радиус «коллизийной зоны»:

$$r = \sqrt{10^{\frac{A}{10}} (r_i^2 + h^2)} - h^2, \quad (6.119)$$

где r_i — радиус окружности, на которой лежит рассматриваемая станция, и подсчитано число станций, попавших в эту зону («конкурентные станции»). Таким образом, из абонентов с одинаковым числом конкурентных станций были образованы слои станций. В табл. 6.9 и 6.10 представлены данные соответственно по количеству слоев станций с одинаковой пропускной способностью (в зависимости от N) и числу станций в различных слоях. В табл. 6.10 число станций в слое показано при условии, что рассматриваемый слой не является последним; иначе число станций в нем равно общему количеству станций N за вычетом числа станций, находящихся во всех остальных слоях.

Таблица 6.9. Число слоев в зависимости от N

Кол-во станций N	1-5	6-7	8-12	13-20	21-28
Число слоев	1	2	3	4	5
Кол-во станций N	29-39	40-51	52-64	65-80	81-96
Число слоев	6	7	8	9	10

Таблица 6.10. Число станций по слоям

Номер слоя	1	2	3	4	5
Число станций в слое	1	1	1	2	2
Номер слоя	6	7	8	9	10
Число станций в слое	2	3	4	4	4

На рис. 6.44 представлен график зависимости суммарной пропускной способности терминальных станций (S_{up}), пропускной способности БС (S_{down}) и полной пропускной способности сети (S) от числа станций N . Как видно из рисунка, с увеличением числа станций пропускная способность БС резко падает, уменьшая тем самым с ростом N и общую производительность сети.

Влияние уровня шума в сети на ее пропускную способность отражено на рис. 6.45.

Рис. 6.44. Пропускная способность в зависимости от числа станций при $BER = 5 \cdot 10^{-5}$ для механизмов: а — базового доступа и б — RTS-CTS

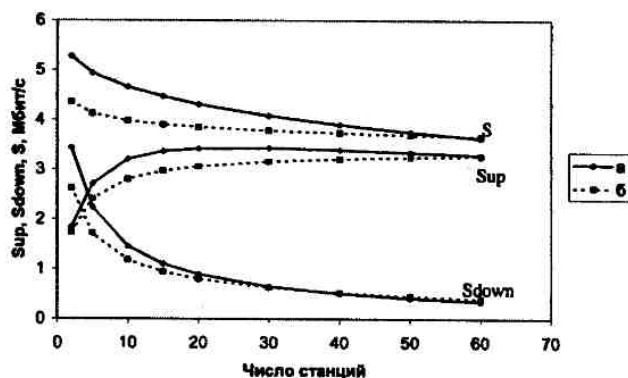
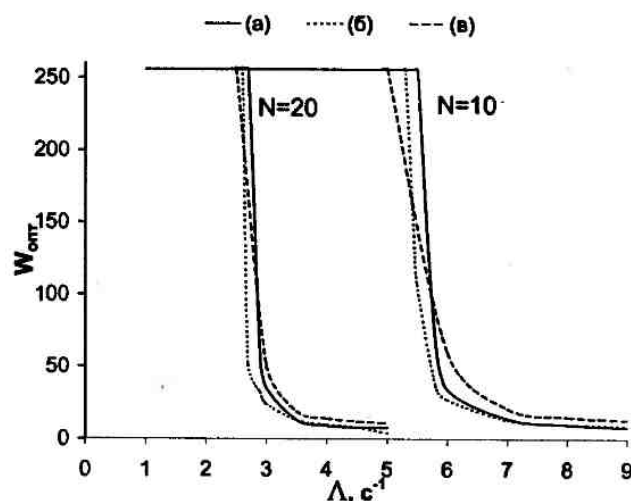


Рис. 6.45. Пропускная способность в зависимости от уровня шума BER для механизмов: а — базового доступа и б — RTS-CTS



В режиме «full-duplex» качественно меняется характер распределения нагрузки на БС и терминальные станции. В этом случае БС будет занимать всю возможную емкость выделенного канала, и ее производительность будет постоянной (при заданном уровне шума) и равной 5,94 Мбит/с при базовом доступе и 4,99 Мбит/с для RTS-CTS механизма. На рис. 6.46 изображена зависимость суммарной пропускной способности терминальных станций (т.е. загруженность в Мбит/с канала «снизу-вверх») такой сети.

Таким образом, в «full-duplex»-сетях, помимо большей пропускной способности по сравнению с сетями с общим каналом передачи, наблюдается значительное увеличение эффективности механизма базового доступа. В итоге (см. рис. 6.47) общая пропускная способность сети возрастает до 10 Мбит/с — верхняя пара кривых, в то время как «half-duplex»-сеть (нижняя пара кривых) способна обеспечить двусторонний обмен данными лишь со скоростью около 5 Мбит/с. В этом смысле режим «full-duplex» действительно оправдывает свое название, превышая в два раза по скорости обмена данными режим «half-duplex».

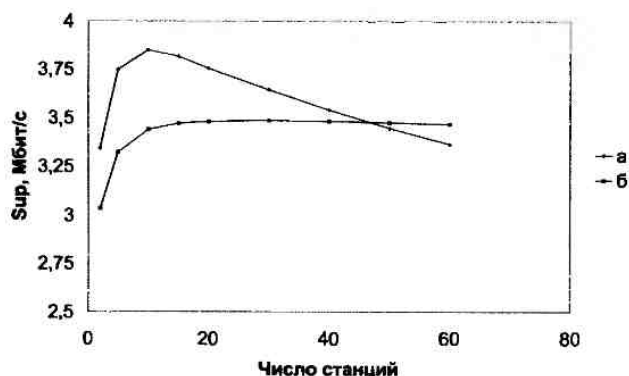


Рис. 6.46. Пропускная способность сети в режиме «full-duplex» в направлении «к БС»

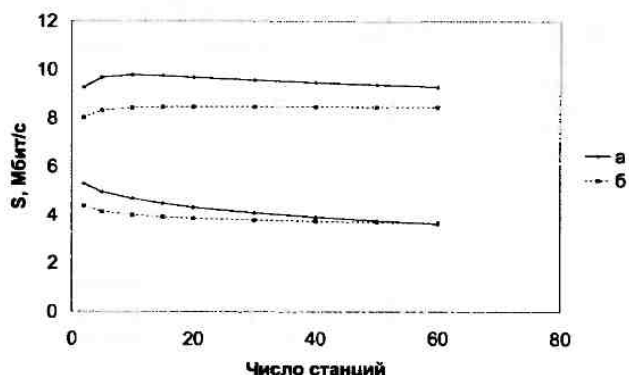


Рис. 6.47. Полная пропускная способность сетей, работающих в режимах «half-duplex» и «full-duplex»

Нахождение оптимальных значений минимального конкурентного окна каждой станции для обеспечения равных условий доступа к предоставляемым сервисам проводилось именно для сетей с двумя частотными каналами. Итак, на рис. 6.48 изображена зависимость емкости канала передачи $S_{up}^{[i]}$, доступной станциям, находящимся на различных расстояниях от БС, до оптимизации и после. Вычисления проводились для $N = 100$ станций при уровне шума $BER = 5 \cdot 10^{-5}$ и использовании механизма доступа RTS-CTS (как более эффективного при таком числе активных станций). Оптимизированная общая пропускная способность уменьшается лишь на 2% по сравнению с первоначальной, но при этом каждая станция обладает равными условиями доступа к каналу передачи. Различие пропускных способностей станций после оптимизации не превышает 1,5%.

Таким образом, в данном разделе была построена простая, но эффективная аналитическая модель для оценки пропускной способности «hot spot»-сети, позволяющая находить оптимальные параметры протокола, реализующие равные условия и качество беспроводного доступа для всех ТС рассмотренной сети.

Оценена производительность сети типа «hot spot», работающей в двух различных режимах — «full-duplex» и «half-duplex». При этом показано, что первый режим работы является более эффективным, чем второй, обеспечивая в два раза большие скорости обмена данными.

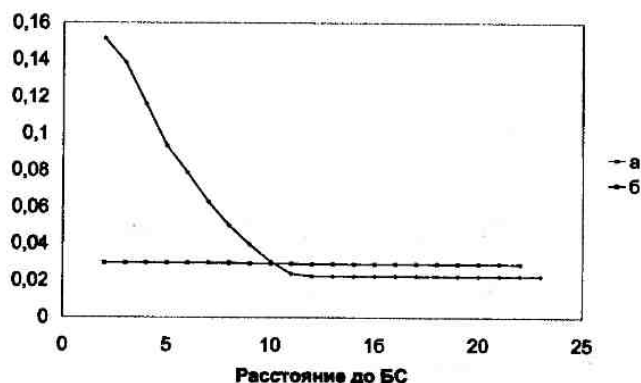


Рис. 6.48. Емкость канала передачи данных S_{up}^i (Мбит/с), приходящаяся на одну терминальную станцию, в зависимости от расстояния до БС: а) до оптимизации, б) после оптимизации

Наконец, на рис. 6.49 показано распределение значений минимального конкурентного окна в зависимости от расстояния до БС. Именно при такой зависимости и достигается равенство пропускных способностей ТС.

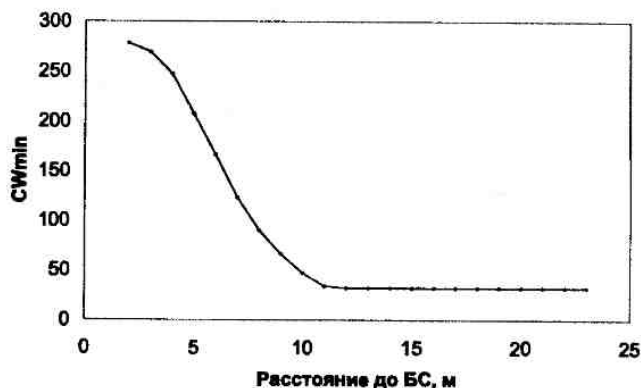


Рис. 6.49. Минимальное значение конкурентного окна в зависимости от расстояния до БС

Найдены оптимальные значения минимального конкурентного окна, гарантирующие равное качество сервиса для всех терминальных станций сети независимо от их расположения относительно БС. При этом функция равного качества обслуживания легко может быть реализована в оборудовании беспроводного доступа, поддерживающем спецификацию IEEE 802.11e. Для этого нужно лишь задать дополнительный тип трафика протокола 802.11e, для которого определить возможность изменения минимального значения конкурентного окна согласно полученной зависимости от расстояния до БС.

ГЛАВА 7

Оценка производительности городских и региональных беспроводных сетей

7.1 Работа протокола IEEE 802.11 в условиях городских и региональных сетей

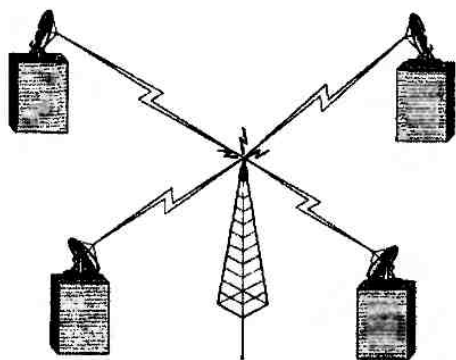


Рис. 7.1. Типовая радиосота

В последние годы беспроводные сети передачи данных становятся одним из основных направлений развития сетевой индустрии. Беспроводная технология позволяет в кратчайшие сроки и с небольшими затратами объединить удаленные локальные сети различных организаций в единую региональную сеть передачи данных, обеспечивая им удаленный высокоскоростной стационарный доступ в Интернет.

Основной структурной единицей региональной беспроводной сети является радиосота (рис. 7.1), имеющая звездообразную структуру: в центре нее помещается базовая станция (БС)

со всенаправленной антенной, на которую сфокусированы антенны оконечных станций (ОС) — радиобриджей между беспроводной сетью и локальными кабельными сетями. В типичных условиях региональной беспроводной сети ОС не имеют радиовидимости друг с другом (т.е. скрыты друг от друга) и вынуждены взаимодействовать через ретрансляционную БС, расположенную на большой высоте (на высотных зданиях, телевышках и т.д.), как это показано на рис. 7.1, и являющуюся точкой доступа во внешнюю сеть.

Примером успешной реализации подобной сети на базе протокола IEEE 802.11 является сеть Radionet [14,25] (рис. 7.2), состоящая из нескольких радиосот, разработанная и реализованная Институтом проблем передачи информации РАН для подключения в Интернет локальных сетей многочисленных организаций науки и образования г. Москвы.

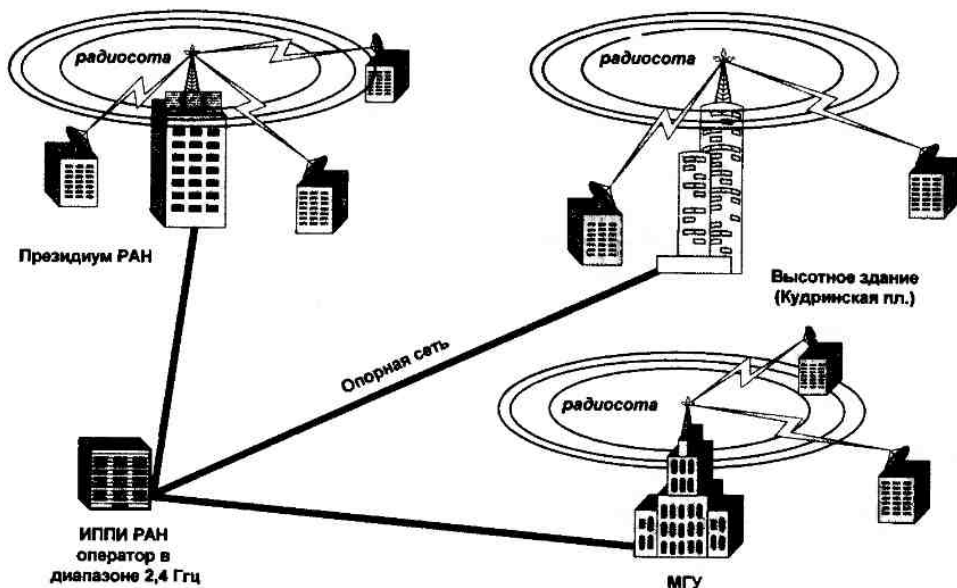


Рис. 7.2. Сеть Radionet

В этом разделе дается краткое описание схемы IEEE 802.11 DCF (см. гл. 6) с обзором лишь тех ее особенностей, которые проявляются в условиях рассматриваемой звездообразной топологии сети со скрытыми ОС при высокой нагрузке на БС. При использовании механизма базового доступа схемы DCF станция, успешно принявшая кадр DATA, содержащий информационный пакет, спустя короткий интервал SIFS немедленно отвечает положительным подтверждением ACK (см. рис. 7.3). Таким образом, попытка передачи пакета считается успешной, если станция, отправившая кадр DATA, в течение тайм-аута EIFS получает корректное подтверждение ACK на этот кадр.

После завершения попытки передачи очередного пакета станция переходит в состояние отсрочки спустя интервал задержки DIFS, если попытка была успешной (т.е. коллизия отсутствовала и все кадры, относящиеся к данному пакету, были переданы корректно, без искажения помехами), или EIFS при неудачной попытке. Заметим, что отсчет времени отсрочки станция ведет только при свободном канале: значение счетчика уменьшается на единицу только в том случае, если в течение всего предшествующего слота канал воспринимался данной станцией как свободный (скрытые станции по-разному воспринимают состояние канала, см. рис. 7.3). Отсчет времени отсрочки прекращается, когда канал становится занят (в восприятии данной станции), и в следующий раз счетчик отложенного времени уменьшится только тогда, когда канал окажется свободен в

течение $\sigma + t_{IF}$, где $t_{IF} = \text{DIFS}$ или $t_{IF} = \text{EIFS}$ в зависимости от того, был ли кадр, принятый данной станцией, последним, соответственно, корректным или нет. При достижении счетчиком отсрочки нулевого значения станция начинает передачу.

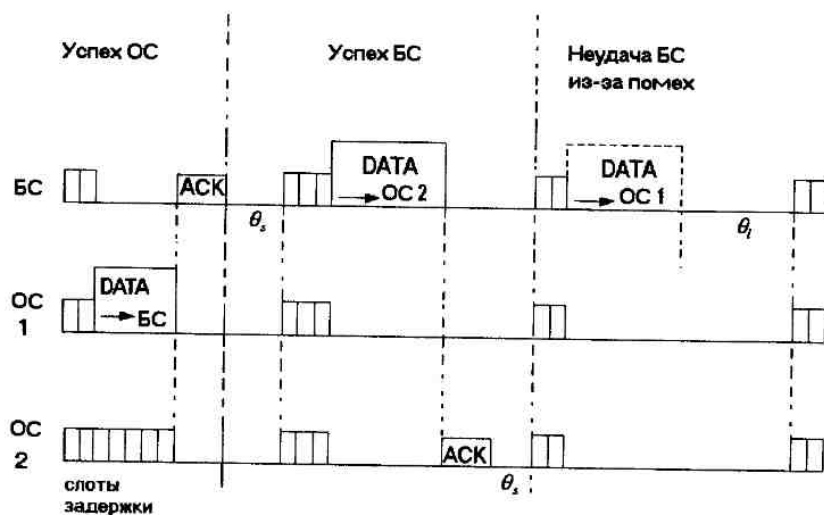


Рис. 7.3. Работа DCF при скрытых ОС ($\theta_l = \text{EIFS}$, $\theta_s = \text{DIFS}$)

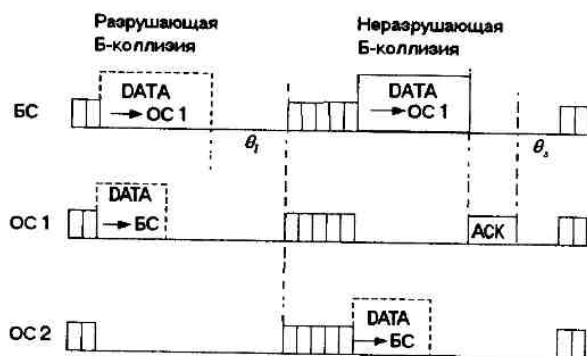


Рис. 7.4. Коллизии BC

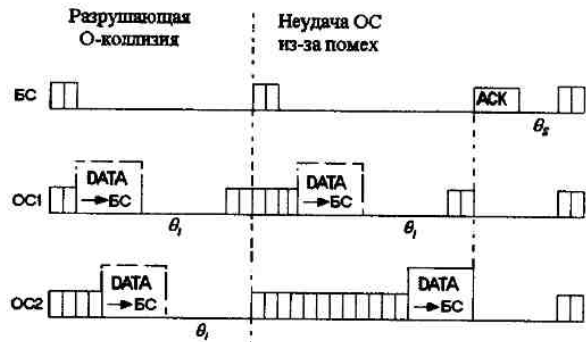
Рассмотрим случаи, когда происходят коллизии, в предположении, что длины всех пакетов, передаваемых от ОС к БС и обратно, равны соответственно τ_e и $\tau_b \gg \tau_e$ байт (данный вид трафика характерен для региональной сети). Кроме того, предположим, что все станции используют только механизм базового доступа. Пусть в некоторый момент t_0 БС, «слышащая» все остальные станции, начинает передачу ОС 1 (см. рис. 7.4). Тогда

коллизия возможна только в случае, если какая-либо ОС также попытается передавать в интервале $(t_0 - \delta, t_0 + \delta)$, где δ — время распространения сигнала, причем передача БС окончится неудачей из-за коллизии, только если в коллизии участвует и ОС 1. Попытки остальных ОС не мешают БС передавать кадр DATA и принимать кадр ACK, так как: 1) станция не прослушивает канал в течение своей передачи; 2) ОС 1 не «слышит» остальные оконечные станции и успешно принимает кадр DATA от БС; 3) $\tau_b > \tau_e$, поэтому коллизия всегда завершается до начала передачи кадра ACK ОС 1. Таким образом, при коллизии, в которой

участвует как БС, так и ОС 1, искажаются все передаваемые кадры (назовем такую коллизию разрушающей Б-коллизией), а если ОС 1 не участвует в коллизии, то данная Б-коллизия является неразрушающей для БС (хотя, конечно, кадры других станций будут искажены).

Перейдем к коллизиям ОС (рис. 7.5). Пусть в момент t_0 начинает передачу ОС 1. Время, необходимое для передачи кадра DATA от ОС, обозначим через t_{de} (при передаче от БС к ОС — t_{db}). Пусть также БС не начнет передачу в $(t_0 - \delta, t_0 + \delta)$. Тогда коллизия произойдет, если какая-либо ОС (ОС 2 на рис. 7.5) также попытается начать передачу в момент $t_1 \in (t_0 - t_{de} - 2\delta - \text{SIFS}, t_0 + t_{de} + 2\delta + \text{SIFS})$, причем при $t_1 \in (t_0 - t_{de}, t_0 + t_{de})$ искажаются оба передаваемых кадра (разрушающая О-коллизия), а при $|t_1 - t_0| \in (t_{de}, t_{de} + 2\delta + \text{SIFS})$ БС успевает успешно принять кадр, прибывший первым, т. е. имеет место неразрушающая О-коллизия. Наконец, после коллизии ОС или искажения помехами кадра DATA, посланного ОС, как эта (или эти) ОС, так и БС не могут начать следующую попытку передачи в течение, по крайней мере, EIFS (будем говорить, что они становятся блокированными на время EIFS), в то время как остальные ОС продолжают отсчет отложенного времени и могут начать передачу.

Рис. 7.5. Неудачные попытки ОС



7.2 Моделирование радиосоты

В этом разделе рассматривается типовая радиосота [24, 160], структура которой упрощенно изображена на рис. 7.6. Радиосота объединяет S оконечных станций, каждая из которых подключает к радиосоте локальную кабельную сеть s ($s = 1, \dots, S$) типа Ethernet, состоящую из N_s терминалов и одного или нескольких серверов. Основываясь на данных из [169], будем учитывать (как и в [24]) только наиболее распространенные информационные взаимодействия типа: запрос к локальному (внутри радиосоты) или внешнему (в глобальной сети) серверу, состоящий из одного пакета, — ответ сервера, состоящий из множества пакетов.

7.2.1 Имитационное моделирование радиосоты

Целью исследования является анализ пропускной способности радиосоты, поэтому прохождение запросов и ответов по локальным кабельным сетям и вне

радиосоты моделируется следующим упрощенным образом. Запрос терминала локальной сети s генерируется в среднем за время λ_s^{-1} с момента приема последнего пакета ответа соответствующей оконечной станцией s и поступает в очередь к этой станции. Извне через БС также поступает пуассоновский поток запросов к локальным серверам с интенсивностью λ_0 , причем число внешних запросов, одновременно обслуживаемых в радиосоте, не превышает N_0 .

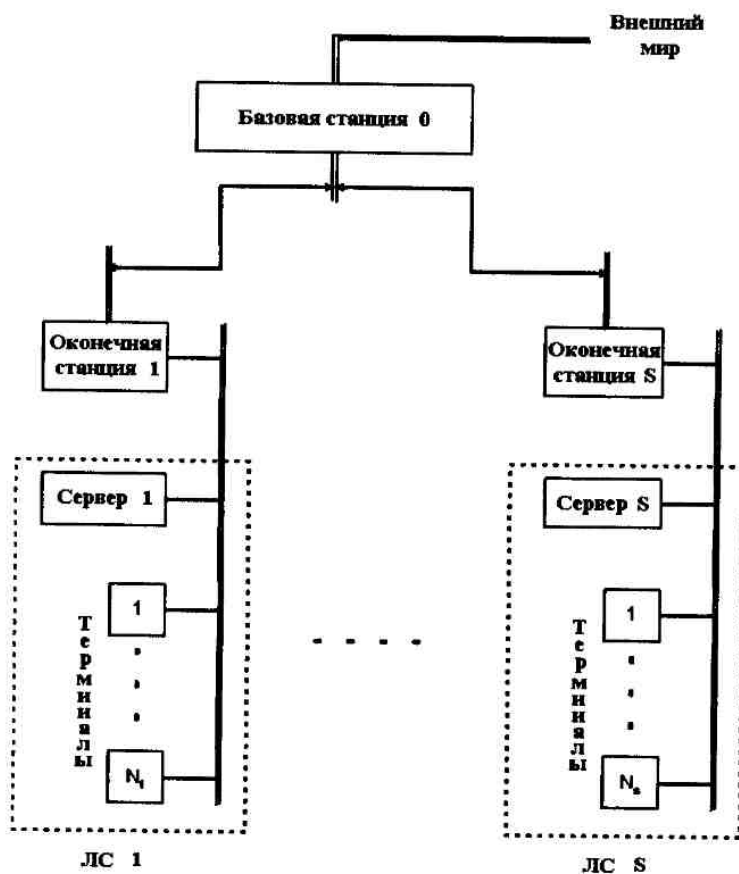


Рис. 7.6. Структура радиосоты с подключенными локальными сетями

После прохождения радиосоты запрос обслуживается в требуемом сервере локальной сети r (при $r > 0$) или «внешнего мира» (при $r = 0$), что определяется маршрутной матрицей $\|p_{sr}\|_{s,r=0,\dots,S}$, где $s = 0$ соответствует внешнему потоку запросов. После обслуживания сервером (в среднем за время μ_{rs}^{-1}) пакеты ответа поступают в очередь к передающей машине станции r (при $r = 0$ — сразу к БС) через случайные интервалы t_r^w (со средним значением $1/\mu_r^w$). Наконец, после передачи последнего пакета ответа через радиосоту к станции-источнику s терминал, ожидавший этого ответа, может генерировать новый запрос (при $s = 0$ освобождается место для нового внешнего запроса).

Времена генерации запросов и обслуживания в серверах (а также интервалы t_r^w) распределены экспоненциально, а число пакетов в ответе сервера r

на запрос из станции s (со средним F_{rs}) имеет геометрическое распределение. Размер пакета ответа L_M предполагается фиксированным и равным 1518 байт (максимальный размер пакета в Ethernet [138]), а размер пакета-запроса — равномерно распределенным на интервале $[0, L_M]$. Наконец, вероятность p_e искажения кадров помехами считается прямо пропорциональной времени передачи кадра. В частности, для кадра DATA с нефрагментированным пакетом ответа $p_e = k_e(t_h + 8L_M/V) = p_e^0$, где $T_M = t_h + L_M/V$ и t_h — времена передачи всего кадра и его заголовка, а V — быстродействие радиосоты в Мбит/с.

Проведенное (на основе сделанных предположений) имитационное моделирование (с помощью языка GPSS [157, 195]) радиосоты с пятью оконечными станциями, результаты которого отчасти приведены в разделе 7.3, потребовало значительных временных затрат даже при небольших количествах терминалов в локальных сетях (N_s) и пакетов (F_{rs}) в файлах-ответах, что объясняется прежде всего существенным дисбалансом временных масштабов: например, интервалы между поступлениями запросов клиентов исчисляются секундами и десятками секунд, а временной масштаб событий, определяемых протоколом радиосоты, измеряется в микросекундах. С ростом же величин N_s и F_{rs} ресурсоемкость имитационного моделирования становится неприемлемой для задач начальных этапов проектирования. Поэтому наиболее эффективным способом определения оптимальной конфигурации радиосоты является ее аналитическое моделирование.

7.2.2 Аналитический метод оценки пропускной способности

Итак, рассмотрим радиосоту на рис. 7.6. Так как число терминалов в каждой локальной сети обычно велико [8, 169], применим вывод из [55] о том, что при оценке пропускных способностей можно заменить конвейерную передачу (через оконечную и базовую станции) файлов-ответов на последовательную. Тогда, пренебрегая интервалами t_r^w между поступлениями последовательных пакетов ответов (эти интервалы достаточно малы и «накладываются» на время ожидания в очереди), получаем замкнутую экспоненциальную модель, изображенную на рис. 7.7.

В этой модели циркулируют заявки $(S + 1)^2$ классов, разбитых на $S + 1$ укрупненных классов. Из IS-станции IS_s^t , моделирующей работу терминалов локальной сети s , исходят заявки класса s_s (при $s = 0$ — это запросы «внешнего мира», генерация которых согласно [14] моделируется одноканальным устройством ОУЗ), которые после прохождения через одноканальные устройства (ОУ i , $i = 0, \dots, S$), моделирующие станции, и обслуживания в требуемом сервере r (также IS-станции IS_r^c) меняют свой класс на s_r . Общее число заявок укрупненного класса s (включающего классы s_r , $r = 0, \dots, S$) фиксировано и равно N_s .

Ненулевые вероятности переходов между узлами модели на рис. 7.7 имеют вид:

для заявок класса s_s при $s > 0$

$$Prob[IS_s^t \rightarrow OY s] = 1, \quad Prob[OY s \rightarrow OY 0] = 1 - p_{s0},$$

$$Prob[OY\ s \rightarrow IS_0^c] = p_{s0},$$

$$Prob[OY\ 0 \rightarrow IS_r^c] = p_{sr}/(1 - p_{s0}), \quad r > 0, \quad r \neq s;$$

для заявок класса s_r при $s > 0$ и $r \neq s$

$$Prob[IS_r^c \rightarrow OY\ r] = 1, \quad Prob[OY\ r \rightarrow OY\ 0] = 1/F_{rs}, \quad r > 0,$$

$$Prob[OY\ 0 \rightarrow IS_s^t] = 1/F_{rs};$$

для заявок класса 0_0

$$Prob[OY3 \rightarrow OY\ 0] = 1, \quad Prob[OY\ 0 \rightarrow IS_r^c] = p_{0r}, \quad r > 0;$$

для заявок класса 0_r при $r > 0$

$$Prob[IS_r^c \rightarrow OY\ r] = 1, \quad Prob[OY\ r \rightarrow OY3] = 1/F_{r0}.$$

В этих выражениях $\sum_{r=0, r \neq s}^S p_{sr} = 1$ для всех $s = 0, \dots, S$.

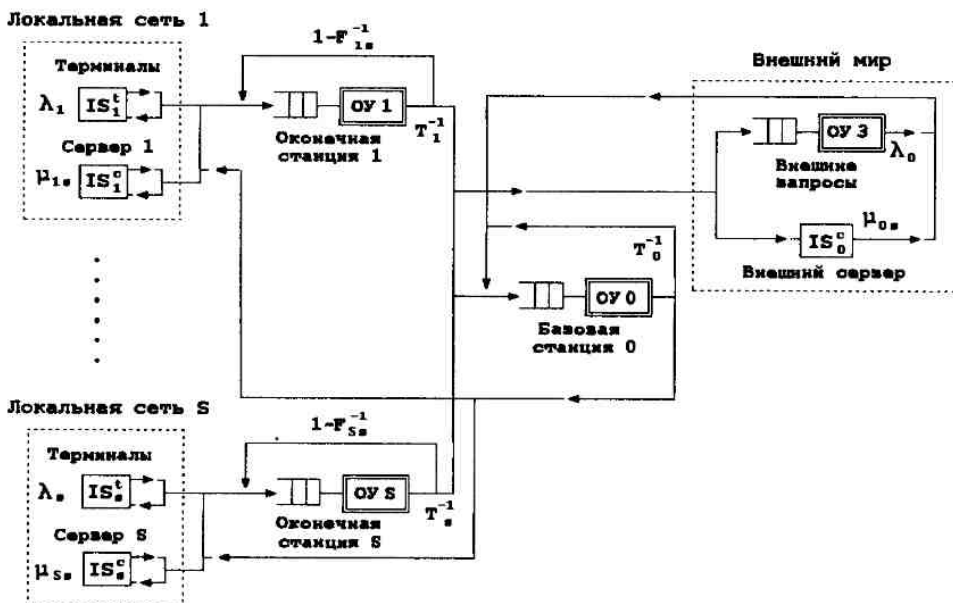


Рис. 7.7. Модель радиосети

Предположив известными средние времена передачи пакета T_s каждой станцией $s = 0, \dots, S$, из этой модели с помощью метода, изложенного в подразделе 3.2.4 монографии [14], находим для каждой станции s : искомые пропускные способности Λ_s по запросам из каждой локальной сети $s = 1, \dots, S$ и по внешним запросам ($s = 0$), а также вероятности ρ_s отсутствия очередей к каждой из станций $s = 0, \dots, S$. При большом числе терминалов N_s точные методы оказываются чрезмерно ресурсоемкими и значения этих показателей целесообразно определять из следующих асимптотических приближений [8]:

$$\Lambda_s = N_s/Q_s, \quad \rho_s = 1 - \sum_{i=0}^S \Lambda_i K_{is},$$

где коэффициенты посещения κ_{is} имеют вид:

$$\kappa_{is} = T_s p_{is} F_{si}, \quad i \neq s; \quad \kappa_{ss} = T_s; \quad (7.1)$$

$$\kappa_{i0} = T_0 \left\{ 1 - p_{i0} + \sum_{j=0, j \neq i}^S p_{ij} F_{ji} \right\}, \quad i > 0, \quad (7.2)$$

а значения Q_s находятся из системы уравнений

$$Q_s = 1/\lambda_s + T_s y_s + \sum_{i=0, i \neq s}^S (p_{si}/\mu_{is} + \kappa_{si} y_i), \quad s > 0, \quad (7.3)$$

$$Q_0 = y_0/\lambda_0 + T_0 y_0 + \sum_{i=1}^S (p_{0i}/\mu_{i0} + \kappa_{0i} y_i), \quad (7.4)$$

$$1 - 1/y_0 - N_0/(\lambda_0 Q_0) = 0, \quad 1 - 1/y_s - \sum_{i=0}^S N_i \kappa_{is}/Q_i = 0, \quad s > 0,$$

с ограничениями $Q_s > 0$, $y_s > 0$ для всех s . Кроме того, из значений Λ_s имеем [24] интенсивности поступления пакетов для передачи на станцию s : $M_s = D_s^0 + D_s^1$, где D_s^0 и D_s^1 — интенсивности поступления соответственно пакетов-запросов и пакетов-ответов, определяемые выражениями

$$D_s^1 = \sum_{i=0, i \neq s}^S F_{si} p_{is} \Lambda_i, \quad D_s^0 = \Lambda_s, \quad s = 1, \dots, S;$$

$$D_0^0 = \Lambda_0 + \sum_{i=1}^S (1 - p_{i0}) \Lambda_i, \quad D_0^1 = \sum_{i=0}^S \sum_{j=1}^S F_{ij} p_{ji} \Lambda_j.$$

Теперь на основе этих показателей будем искать значения T_s . Предположим:

- 1) механизм слежения за эфиром абсолютно надежен;
- 2) длина всех пакетов одинакова, так что каждый из них состоит из $n_f = \lceil L_M/L_f \rceil$ фрагментов, а время передачи любого кадра DATA t_{DATA} и вероятность его искажения помехами соответственно равны

$$t_{\text{DATA}} = t_h + L_M/(n_f^d V) \quad \text{и} \quad p_e^f = p_e^0 t_{\text{DATA}}/T_M; \quad (7.5)$$

- 3) для передачи всех пакетов используются кадры RTS и CTS (что рекомендуется стандартом IEEE 802.11 в случае «скрытых» станций), так что отсутствуют коллизии кадров DATA;

- 4) число попыток передачи пакета не ограничено;

- 5) при подсчете тайм-аутов и t_{nav} используются (см. раздел 6.6) реальные значения τ_{as} времени распространения сигнала между оконечной станцией s и базовой станцией.

Кроме того, не будем учитывать потери от коллизирующих кадров RTS в части бесполезного занятия эфира. Обозначим: $\tau_0 = \text{SIFS}$, $\tau_g = \text{DIFS}$, $\tau_b = \text{EIFS}$.

Передача пакета может начаться только в те моменты времени, когда эфир свободен и с момента занятости эфира прошло, по крайней мере, время τ_g или τ_b . Пренебрегая потерями от коллизирующих кадров RTS в части бесполезного за-

нения эфира (включая интервалы τ_b), долю U этих моментов времени можно приближенно оценить выражением $U = 1 - \sum_{s=0}^S U_s$, где доля моментов U_s , $s = 1, \dots, S$, когда эфир занят передачей пакета от станции s , определяется по формулам:

$$\begin{aligned} U_s &= M_s \{ (T_s^0 + \tau_g) + (\hat{U}_s + T_M - t_h) / (1 - p_e^m) \}, \\ \hat{U}_s &= n_f [t_h + t_{ACK} + \tau_s^r + p_e^m (T_s^0 + \tau_b)]. \end{aligned} \quad (7.6)$$

В этих формулах

$T_s^0 = t_{RTS} + t_{CTS} + 2\tau_{as} + \tau_0$, $\tau_s^r = 2(\tau_0 + \tau_{as})$, $p_e^m = 1 - (1 - p_e^f)(1 - p_e^a)$; $p_e^f = p_e^0(t_{RTS} + t_{CTS})/T_M$ и $p_e^a = p_e^0 t_{ACK}/T_M$ — вероятности искажения помехами соответственно одного из последовательных кадров RTS и CTS и кадра ACK.

Вероятность неуспешной попытки передачи пакета от станции s равна:

$$P_{hs} = 1 - (1 - p_e^r)(1 - P_{cs})(1 - p_e^m), \quad (7.7)$$

где P_{cs} — вероятность коллизии при передаче RTS. Найдем эту величину. Пусть передача RTS от оконечной станции s началась в момент t_0 . Тогда коллизия с RTS от другой оконечной станции i произойдет в случае, если та начала передачу в интервале $(t_0 - t_{RTS} - \tau_0 - \tau_{as} - \tau_{ai}, t_0 + t_{RTS} + \tau_{as} - \tau_{ai})$; для $i = 0$ (коллизия с базовой станцией) этот интервал равен $(t_0 - \tau_{as}, t_0 + \tau_{as})$. Следовательно, предполагая (здесь и далее) пуассоновский характер распределения моментов начала передачи на интервалах свободного эфира, имеем:

$$P_{cs} = 1 - \exp \left\{ -\frac{2M_0\eta\tau_{as}}{(1 - P_{h0})U} \right\} \exp \left\{ -t_s^{coll} \sum_{i=1, i \neq s}^S \frac{M_i\eta}{(1 - P_{hi})U} \right\}, \quad s = 1, \dots, S, \quad (7.8)$$

$$P_{c0} = 1 - \exp \left\{ -2 \sum_{i=1}^S \frac{M_i\eta\tau_{ai}}{(1 - P_{hi})U} \right\}, \quad (7.9)$$

где

$$t_s^{coll} = 2(t_{RTS} + \tau_{as}) + \tau_0, \quad \eta = 1 + p_e^m(n_f - 1).$$

Решая систему уравнений (7.7)–(7.9), находим P_{hs} .

Определим среднее время пребывания станции s в состоянии отсрочки в расчете на один пакет, включая время возможного ожидания освобождения канала. Рассмотрим *случай 1* передачи пакета, поступившего в непустую очередь. Учитывая, что очередь может быть пуста только при поступлении пакета-запроса или первого из пакетов-ответов, находим вероятность этого случая: $\xi_{s1} = 1 - \rho_s q_s^1$, где

$$q_0^1 = \left(D_0^0 + \sum_{i=0}^S \sum_{j=1}^S p_{ji} \Lambda_j \right) / M_s, \quad q_s^1 = \left(D_s^0 + \sum_{i=0, i \neq s}^S p_{is} \Lambda_i \right) / M_s, \quad s > 0.$$

Исходя из (6.2), среднее время отсрочки определяется по формуле

$$T_s^b = A_{s1} + \frac{\sigma}{2\psi_s} \sum_{j=0}^{\infty} \varphi_j^{(s)} \bar{W}_j \quad (7.10)$$

с коэффициентом $A_{s1} = 0$, где $\bar{W}_j$ — суммарное время отсрочки при j неудачных попытках передачи пакета, т. е.

$$\bar{W}_j = \sum_{i=n_c^{\min}}^{n_c^{\min}+j} (2^i - 1) \text{ при } j \leq m = n_c^{\max} - n_c^{\min} \text{ и}$$

$$\bar{W}_j = \bar{W}_m + (2^{n_c^{\max}} - 1)(j - m) \text{ при } j > m,$$

где

$$2^{n_c^{\min}} = W_0 \text{ и } 2^{n_c^{\max}} = W_m,$$

а вероятность того, что произошло ровно j неудачных попыток, равна

$$\varphi_j^{(s)} = \sum_{k=1}^{\min\{j+1, n_f\}} B_k \binom{j}{k-1} P_{hs}^{j-k+1},$$

где

$$B_k = (1 - P_{hs})^k (p_e^m)^{k-1} (1 - p_e^m)^{n_f-k};$$

наконец, делитель $\psi_s = 1 - \sum_{i=0, i \neq s}^S U_i / (1 - U_s)$ в (7.10) отражает тот факт, что отсчет времени отсрочки ведется только при свободном эфире. Для удобства вычисления формулу (7.10) можно представить в виде:

$$T_s^b = A_{s1} + \frac{t_{slot}}{2\psi_s} \left\{ \sum_{j=0}^J \varphi_j^{(s)} \bar{W}_j + \sum_{k=1}^{n_f} \left[B_k^0 \frac{d^{k-1} f(P_{hs})}{dP_{hs}} + B_k^1 \frac{d^k f(P_{hs})}{dP_{hs}} \right] \right\}, \quad (7.11)$$

где

$$J = \max\{n_f, m\}, \quad f(P_{hs}) = P_{hs}^{J+1} / (1 - P_{hs}),$$

$$B_k^0 = [\bar{W}_J + (2^{n_c^{\max}} - 1)(J - k + 1)] B_k, \quad B_k^1 = (2^{n_c^{\max}} - 1) B_k.$$

Рассмотрим теперь различные случаи поступления пакета в пустую очередь. Формулы (7.10) и (7.11) остаются справедливыми и в этих случаях, но с заменой коэффициента A_{s1} на A_{si} , где i — номер случая.

Случай 2: станция s находится в состоянии отсрочки (включая интервал задержки) после своей последней передачи. Вероятность этого случая: $\xi_{s2} = \rho_s q_s^1 \nu_s$, где

$$\nu_s = 1 - \exp\{-M_s \Delta_{s2}^b\}, \quad \Delta_b = \tau_g + t_{slot}(2^{n_c^{\min}} - 1) / (2\psi_s),$$

а коэффициент: $A_{s2} = -\Delta_b / 2$. В остальных случаях станция s не находится в состоянии отсрочки.

Случай 3: при поступлении пакета либо эфир занят, либо не истек интервал задержки после передачи другой станции. Вероятность:

$$\xi_{s3} = \rho_s q_s^1 (1 - \nu_s) (1 - \psi_s),$$

коэффициент:

$$A_{s3} = [T_s^0 + T_M - t_h + n_f^d(t_h + t_{ACK} + \tau_s^r + 2\tau_{as}) + \tau_g] / 2 - \tau_g.$$

Случай 4: при поступлении пакета эфир свободен и истек интервал задержки. Вероятность: $\xi_{s4} = 1 - \sum_{j=1}^3 \xi_{sj}$, коэффициент: $A_{s4} = -\Delta_b$.

Суммируя (с соответствующими вероятностями ξ_{si}) средние времена отсрочки T_s^b для всех случаев и учитывая (из (7.6)), что фактическое среднее время

занятости эфира при передаче одного пакета станцией s примерно равно

$$\frac{U_s}{M_s} + \Delta_s^{coll}, \text{ где } \Delta_s^{coll} = \eta \left[\frac{1}{1 - P_{hs}} - \frac{1}{1 - p_e^m} \right] (t_{RTS} + \tau_{as} + \tau_b)$$

— потери станции s из-за коллизий, находим искомое среднее время передачи пакета:

$$T_s = \frac{U_s}{M_s} + \Delta_s^{coll} + \hat{T}_s^b,$$

где $\hat{T}_s^b$ определяется так же, как T_s^b , но с заменой A_{s1} на усредненный коэффициент $\bar{A}_s = \sum_{j=1}^4 \xi_{sj} A_{sj}$. Подставляя значения T_s в (7.1)–(7.4), получаем искомые пропускные способности.

7.2.3 Метод оценки пропускной способности при технологии FHSS

При реализации радиосоты на базе технологии FHSS все станции работают на одной частоте в течение фиксированного временного интервала T_{dwell} , по окончании которого происходит их переключение (в течение короткого интервала t_{cf}) на другую частоту; далее процесс повторяется. При переключении частот теряется контроль за состоянием эфира. Поэтому станция, желающая передать пакет (или фрагмент пакета), воздерживается от передачи, если предполагаемое время передачи (рассчитываемое исходя из максимального времени распространения сигнала $\tau_a^{\max}$) превышает время, оставшееся до конца интервала работы на текущей частоте, и переходит в отложенное состояние, не изменяя текущее значение счетчика неудач n_r [222]. Назовем такую попытку передачи отложенной. При выходе из состояния отсрочки станция возвращается в него, если указанное условие допустимости передачи не выполняется.

Пусть передается кадр DATA с фрагментом, не являющимся последним, и предполагаемое время передачи последующего фрагмента не укладывается в оставшееся время работы на текущей частоте. Тогда, во-первых, значение t_{nav} (в поле предполагаемой длительности передачи) для передающегося кадра DATA равно времени передачи АСК, а t_{nav} для ответного кадра АСК обнуляется, и, во-вторых, после получения этого ответного кадра АСК станция переходит в состояние отсрочки, не начиная передачу последующего фрагмента. Таким образом, решение о том, передавать фрагмент $i > 1$ в текущем интервале T_{dwell} или нет, принимается в момент начала передачи кадра DATA предыдущего фрагмента $i - 1$.

Рассмотрим сначала случай, когда фрагментация пакетов не используется. При этом среднее значение предполагаемого времени передачи для станции s равно

$$T_s^p = (D_s^0/2 + D_s^1)(T_M - t_h) + t_h + t_{RTS} + t_{CTS} + t_{ACK} + 3\tau_0 + 4\tau_a^{\max}.$$

Тогда среднее время от отложенной попытки передачи до конца интервала работы на текущей частоте (включая интервал переключения), когда станции запрещено передавать, равно $T_s^p/2 + t_{cf}$.

Другим фактором снижения производительности по сравнению с режимом без частотных скачков является тот факт, что после отложенной попытки станция переходит в состояние отсрочки и поэтому делает попытку передачи не в начале следующего интервала T_{dwell} , а спустя некоторое время. Однако в случае, когда последняя успешная передача заканчивается в момент времени, достаточно близкий к концу интервала T_{dwell} , во-первых, следующая попытка передачи не будет отложенной и, во-вторых, время нахождения в фазе отсрочки между этими попытками будет существенно меньше, чем в режиме без частотных скачков (ввиду малой загрузки эфира в конце интервала T_{dwell}).

Поэтому можно считать (и это подтверждается имитационным моделированием), что влияние двух последних факторов взаимно компенсируется и режим с частотными скачками отличается от режима без них только периодами «молчания» со средней продолжительностью $T_s^p/2 + t_{cf}$ в конце каждого интервала T_{dwell} .

Следовательно, исключая эти периоды, можно предположить, что в оставшееся время станция s работает так же, как аналогичная станция без частотных скачков, обеспечивающая интенсивности потоков запросов $\hat{D}_s^0 = D_s^0/(1 - \delta_s)$ и ответов $\hat{D}_s^1 = D_s^1/(1 - \delta_s)$, где $\delta_s = (T_s^p/2 + t_{cf})/(T_{dwell} + t_{cf})$. Таким образом, среднее время передачи пакета T_s будем рассчитывать по формулам раздела 7.2.2 с заменой интенсивностей потоков запросов D_s^0 и ответов D_s^1 на $\hat{D}_s^0$ и $\hat{D}_s^1$. В свою очередь, при определении пропускных способностей Λ_s и вероятностей отсутствия очередей ρ_s , $s = 0, \dots, S$, в формулы (7.1)–(7.4) вместо T_s следует подставить $\hat{T}_s = T_s/(1 - \delta_s)$.

Перейдем к случаю фрагментированных пакетов и разделим попытки передачи фрагментов на две категории: категория (а), когда либо номер фрагмента $i = 1$, либо предыдущая попытка передачи не была успешной; и категория (б), когда имеет место первая попытка передачи фрагмента $i > 1$. Обозначим время, требуемое для передачи кадра DATA с фрагментом i , через $t_{DATA}^{(i)}$. Тогда попытка передачи фрагмента i будет отложена, если время до конца интервала работы на текущей частоте меньше либо (при категории (а))

$$t_s^{fa} = t_{DATA}^{(i)} + t_{RTS} + t_{CTS} + t_{ACK} + 3\tau_0 + 4\tau_a^{\max},$$

либо (при категории (б))

$$t_s^{fb} = t_{DATA}^{(i)} + t_{ACK} + \tau_0 + 2(2\tau_a^{\max} - \tau_{as}).$$

Будем считать, что среднее время t_{DATA} , требуемое для передачи кадра DATA, определяется (7.5), и обозначим средние значения t_s^{fa} и t_s^{fb} через T_s^{fa} и T_s^{fb} , которые определяются теми же формулами с заменой $t_{DATA}^{(i)}$ на t_{DATA} .

Попытка категории (а) происходит с вероятностью $1 - p_s^{fb}$, где $p_s^{fb} = (1 - p_e^m)(1 - n_{fs}^{-1})$ — вероятность попытки категории (б), и выполняется в основном после пребывания станции s в состоянии отсрочки. Поэтому для категории (а), как и в случае нефраgmentированных пакетов, можно считать, что эффект переключения частот сводится к периодам «молчания» со средней продолжительностью $T_s^{fa}/2 + t_{cf}$ в конце каждого интервала T_{dwell} .

Попытка категории (б) выполняется сразу (через τ_0) по окончании предыдущей попытки без пребывания станции в состоянии отсрочки. Поэтому для этой

категории необходимо учесть, наряду с периодом «молчания» со средней продолжительностью $T_s^{db} = T_s^{fb}/2 + t_{cf}$, среднее время Δ_s^{db} нахождения в фазе отсрочки в начале следующего интервала T_{dwell} после того, как данная попытка была отложена. Для нахождения Δ_s^{db} учтем следующие факты: 1) среднее время от отложенной попытки передачи до конца интервала работы на текущей частоте (включая интервал переключения) равно T_s^{db} ; 2) средняя продолжительность фазы отсрочки равна $T_b^{\min} = \tau_g + (2^{n_c^{\min}} - 1)t_{slot}/2$. Предположим также, что случайное время от момента начала фазы отсрочки, переходящей на следующий интервал T_{dwell} , до конца текущего интервала T_{dwell} распределено равномерно между 0 и $\min(T_b^{\min}, T_s^{db})$. Тогда в период $(0, \Delta_s^{db})$, отсчитываемый с начала следующего интервала T_{dwell} , усредненная вероятность свободного (с точки зрения данной станции s) эфира ψ_s^d (величина, аналогичная ψ_s) примерно равна

$$\psi_s^d = 1 - \sum_{i=0, i \neq s}^S \frac{U_i(1 - 3p_i^{fb}/4)}{1 - U_s(1 - 3p_s^{fb}/4)},$$

откуда получаем следующее приближение:

$$\Delta_s^{db} = [T_b^{\min} - \min(T_b^{\min}, T_s^{db})/2]/\psi_s^d.$$

Таким образом, в случае фрагментированных пакетов величина δ_s , используемая (см. выше) для модификации средних времен передачи пакета T_s и интенсивностей потоков запросов D_s^0 и ответов D_s^1 , определяется по формуле

$$\delta_s = [(p_s^{fa}T_s^{fa} + p_s^{fb}T_s^{fb})/2 + p_s^{fb}\Delta_s^{db} + t_{cf}]/(T_{dwell} + t_{cf}).$$

Вычисляя с помощью этой величины значения $\hat{D}_s^0$, $\hat{D}_s^1$ и $\hat{T}_s$, $s = 0, \dots, S$, и используя их вместо D_s^0 , D_s^1 и T_s , определяем искомые пропускные способности Λ_s .

7.3 Численные результаты исследования городской радиосоты

Рассмотрим радиосоту, объединяющую $S = 5$ оконечных станций со статистически однородными локальными сетями, т. е. для каждого $s = 1, \dots, S$:

$$N_s = N/S, \quad \lambda_s = \lambda, \quad p_{s0} = p_0, \quad p_{0s} = 1/S, \quad \tau_{as} = 0,000017 \text{ с}, \\ p_{si} = (1 - p_0)/(S - 1) \text{ при } i > 0 \text{ и } i \neq s$$

(значение τ_{as} соответствует расстоянию 5 км). Кроме того, средние количества пакетов-ответов $F_{rs} = F$ для всех серверов r ; после обслуживания любого запроса в любом из локальных серверов r с одинаковой интенсивностью $\mu_{rs} = \mu_1 = 2 \text{ с}^{-1}$ пакеты ответа поступают в среднем через 0,002 с; при обслуживании во «внешнем мире» эти величины равны соответственно $\mu_0 = 0,2 \text{ с}^{-1}$ и 0,02 с. Размер слота выбран из расчета максимально возможного расстояния между станциями, равного 50 км, а минимум длины пакета $L_{\min} = 300$ байт.

Начнем со случая, когда радиосота реализована по технологии FHSS с быстрой скоростью $V = 2$ Мбит/с, достаточно большим временем T_{dwell} и фрагментация не используется, т. е. $L_f > 1518$ байт. Соответствующие параметры протокола

Radio-Ethernet, полученные на основе данных из [222], приведены в табл. 7.1 (временные параметры даны в микросекундах). Вероятность помех p_e^0 при передаче пакета равна $p_e^0 = 0,1$. На рис. 7.8 показаны зависимости суммарной локальной пропускной способности $\Lambda = \sum_{s=1}^S \Lambda_s$ этой радиосоты от среднего времени генерации запроса $T_g = \lambda^{-1}$.

Таблица 7.1. Параметры моделируемой радиосоты: технология FHSS

σ	$\tau_a^{\max}$	τ_0	τ_g	τ_b	t_{RTS}
216	167	28	460	728	208
$t_{\text{CTS}} = t_{\text{ACK}}$	T_M	t_h	$n_c^{\min}$	$n_c^{\max}$	
184	6072	264	4	10	

Эти зависимости получены при $N = 50$, $N_0 = 100$ и $F = 10$ с помощью имитационного моделирования (сплошные линии) и аналитических методов (штриховые линии) для следующих случаев, различающихся вероятностями внешнего доступа p_0 и интенсивности потока внешних запросов λ_0 .

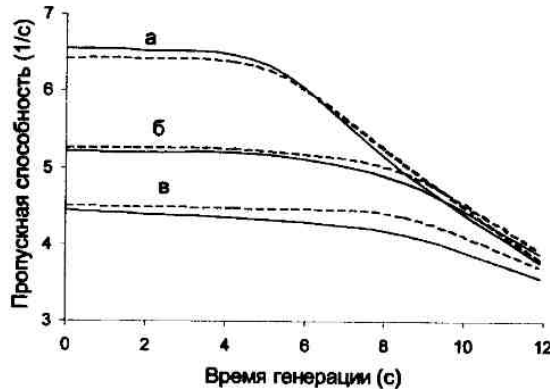


Рис. 7.8. Зависимость суммарной пропускной способности по локальным запросам от среднего времени их генерации для разных случаев: а — $p_0 = 0,5$ и $\lambda_0 = 1 \text{ с}^{-1}$; б — $p_0 = 0,9$ и $\lambda_0 = 1 \text{ с}^{-1}$; в — $p_0 = 0,5$ и $\lambda_0 = 4 \text{ с}^{-1}$

Видно, что при уменьшении T_g кривые входят в область насыщения, т. е. пропускная способность полностью определяется интенсивностью обслуживания в «узком месте», которым является базовая станция (кривые а и б), а в случае в — все станции. При больших T_g пропускная способность примерно равна своему максимальному значению $\Lambda^{\max}$, достижимому при данном T_g и определяемому (из расчета нулевой задержки передачи) из формулы:

$$1/\Lambda^{\max} = N/[T_g + p_0/\mu_0 + (1 - p_0)/\mu_1].$$

Поэтому используем в качестве показателя эффективности работы радиосоты ее коэффициент полезного действия (КПД), равный проценту, который Λ составляет от $\Lambda^{\max}$. Графики зависимости КПД от T_g показаны на рис. 7.9 для случаев а и б (график для случая в не приведен ввиду его практического совпадения со случаем б).



Рис. 7.9. Зависимость КПД сети от среднего времени генерации локальных запросов

Как видно из рис. 7.8 и 7.9, погрешность разработанных приближений нигде не превышает 5%, поэтому данный метод вполне применим для оценки пропускной способности радиосети, причем в отличие от имитационного моделирования может эффективно использоваться при решении поисковых задач проектирования сетей с сотнями и тысячами клиентов. В качестве примера найдем область значений интенсивностей λ и λ_0 , где КПД сети не меньше 80%. При этом установим следующие (более реальные) параметры сети: $F = 50$, $p_0 = 0,5$, $N_0 = 500$, а N принимает два значения: 250 и 500. При имитационном моделировании на решение этой задачи потребовались бы недели, в то время как разработанный метод позволил в течение часа получить следующие результаты (рис. 7.10). Кривые на этом рисунке ограничивают сверху искомую область значений.



Рис. 7.10. Кривые ограничения нагрузки

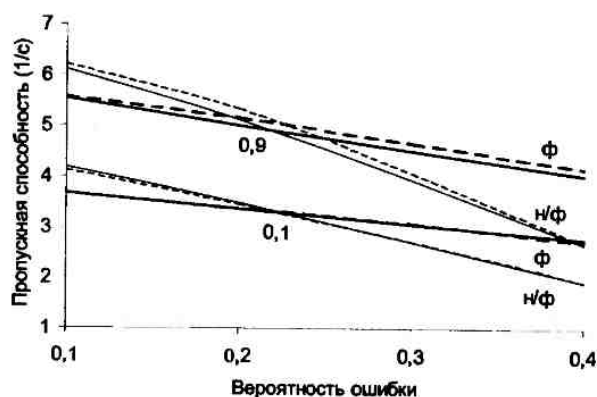
Перейдем к исследованию эффективности использования фрагментации. Рассмотрим случай, когда радиосота реализована по технологии DSSS с быстродействиюм $V = 2$ Мбит/с и $L_f = 400$ байт. Соответствующие параметры протокола [222] даны в табл. 7.2. Кроме того, $N = 50$, $N_0 = 100$, $\lambda = 0,5 \text{ с}^{-1}$, $\lambda_0 = 1 \text{ с}^{-1}$. На рис. 7.11 показаны графики суммарной локальной пропускной способности $\Lambda = \sum_{s=1}^S \Lambda_s$ этой радиосоты в зависимости от интенсивности помех, характеризующей вероятностью ошибки p_e^0 из-за помех при передаче кадра DATA с нефрагментированным пакетом ответа. Эти графики, образующие два семейства, отличающиеся вероятностью внешнего доступа p_0 (0,1 и 0,9), получены

при фрагментированной (толстые линии с меткой «ф») и нефрагментированной (тонкие линии с меткой «н/ф») передаче пакетов с помощью имитационного моделирования (штриховые линии) и разработанных приближений (сплошные линии). Видно, что при малых помехах эффект от использования фрагментации, определяемый разностью значений пропускной способности при фрагментации и без нее, с увеличением интенсивности помех монотонно растет от отрицательных значений (при $p_e^0 < 0,23$) до существенно положительных: в частности, при $p_e^0 = 0,4$ применение фрагментации повышает пропускную способность почти в 1,5 раза. Следует отметить, что эта картина (в том числе и граница эффективности применения фрагментации) качественно не меняется с изменением вероятности внешнего доступа p_0 . Из сравнения графиков, изображенных сплошными и штриховыми линиями, видно, что погрешность разработанных приближений нигде не превышает 5%.

Таблица 7.2. Параметры моделируемой радиосоты: технология DSSS

σ	$\tau_a^{\max}$	τ_0	τ_g	τ_b	t_{RTS}
192	167	10	384	698	272
$t_{CTS} = t_{ACK}$		T_M	t_h	$n_c^{\min}$	$n_c^{\max}$
248		6360	288	5	10

Рис. 7.11. Зависимость суммарной пропускной способности по локальным запросам от вероятности ошибки p_e^0 при технологии DSSS с фрагментацией (ф) и без фрагментации (н/ф) при вероятности внешнего доступа (p_0) 0,1 и 0,9



Численные результаты, изображенные на рис. 7.11, получены в случае высокой нагрузки на радиосоту. Будем равномерно снижать (в K_l раз) интенсивности λ_s для всех $s = 0, \dots, S$ и посмотрим, как это влияет на эффект от фрагментации при фиксированной вероятности ошибки $p_e^0 = 0,33$. Результаты этого исследования, приведенные на рис. 7.12, показывают в целом немонотонность зависимостей суммарной локальной пропускной способности Λ от коэффициента снижения нагрузки K_l . Рост K_l вначале (кроме случая фрагментированной передачи при $p_0 = 0,9$) увеличивает Λ за счет снижения интенсивности внешнего потока. При дальнейшем снижении нагрузки пропускная способность начинает также снижаться и становится примерно равной своему максимальному значению $\Lambda^{\max}$, достижимому при данных λ_s , за счет чего эффект от фрагментации стремится

к 0 (см. рис. 7.12). Таким образом, в случае малой нагрузки пропускные способности при фрагментированной и нефрагментированной передачах практически совпадают.

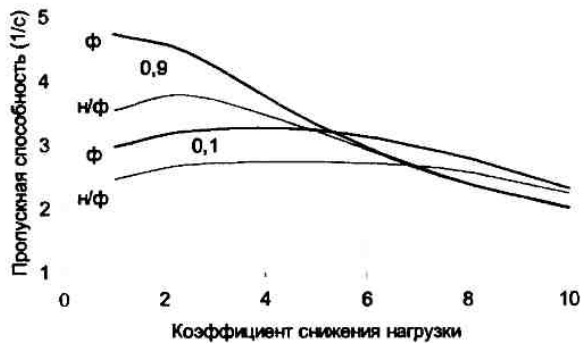


Рис. 7.12. Зависимость суммарной пропускной способности по локальным запросам от коэффициента снижения нагрузки K_l . Обозначения: см. рис. 7.11

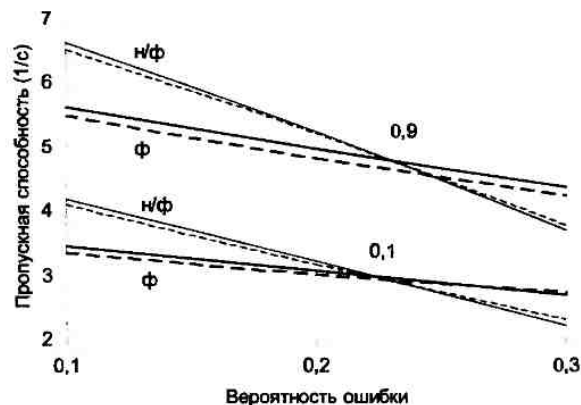


Рис. 7.13. Зависимость пропускной способности от вероятности ошибки p_e^0 при технологии FHSS. Обозначения: см. рис. 7.11

Таблица 7.3. Параметры моделируемой радиосоты: реализация BreezeNet

σ	τ_0	τ_g	τ_b	T_M	t_h
216	28	460	728	4240	192
$t_{CTS} = t_{ACK}$		t_{RTS}	T_{dwell}	t_{cf}	$n_c^{\min}$
165		181	32000	220	4

Наконец, рассмотрим случай, когда радиосота реализована с помощью радиобриджей BreezeNet [69], обеспечивающих максимальное быстродействие $V = 3$ Мбит/с и работающих по технологии FHSS с ограниченным временем T_{dwell} . Измененные (согласно [69, 222]) по сравнению с табл. 7.2 параметры протокола даны в табл. 7.3. На рис. 7.13 показаны (с теми же условными обозначениями, что и на рис. 7.11) графики зависимости суммарной локальной пропускной способности Λ этой радиосоты от интенсивности помех (т. е. от вероятности ошибки p_e^0) для случаев фрагментированной и нефрагментированной передачи при различных вероятностях внешнего доступа p_0 (0,1 и 0,9). Эти графики получены при

значениях $\lambda_s = 1$, $s = \overline{1, S}$, и $\lambda_0 = 4 \text{ с}^{-1}$. Видно, что, во-первых, погрешность разработанных приближений мала и, во-вторых, графики на рис. 7.13 качественно не отличаются от графиков на рис. 7.11, включая практическое совпадение границы эффективности применения фрагментации. Следовательно:

- выводы о зависимости эффективности фрагментации от интенсивности помех и нагрузки, полученные для технологии DSSS, остаются справедливыми и в этом случае;
- при фрагментации положительный эффект, достигаемый за счет увеличения битов информации, передаваемых за период непрерывной работы на одной частоте, практически компенсируется отрицательным эффектом от пребывания станции в состоянии отсрочки (при переключении частот) между передачами фрагментов одного и того же пакета.

7.4 Оценка максимальной производительности беспроводного доступа в Интернет

В предшествующих аналитических работах (см., например, [130, 137, 212]) оценка максимальной пропускной способности, обеспечиваемой протоколом IEEE 802.11, проводилась в основном лишь для локальных сетей при допущениях о постоянном наличии непустых очередей на всех станциях сети и об отсутствии «скрытых» станций. Как показали результаты длительных измерительных экспериментальных исследований (см. раздел 7.7) сети Radionet, в типовой радиосоте (рис. 7.1) преобладает трафик «сверху вниз», т. е. от БС к ОС, так как в основном требуется информация, получаемая из мировой сети Интернет через БС, причем основная доля беспроводного трафика связана с передачей больших файлов посредством TCP-соединений (FTP-подобный трафик).

В связи с этим допущение о постоянном наличии непустой очереди пакетов (т. е. о насыщении этой очереди), готовых к передаче, применимо только для БС, а трафик «снизу вверх» образуется TCP-подтверждениями малой длины на TCP/IP-пакеты, полученные от БС. Более конкретно в данном разделе рассматривается ситуация, когда в течение длительного времени существует множество стабильных TCP-соединений между пользователями локальных сетей, интегрированных в рассматриваемую радиосоту, и внешним миром, причем предполагается, что

- TCP-соединения равномерно распределены по всем ОС, т. е. можно считать, что передаваемый TCP/IP-пакет с равной вероятностью, не зависящей от предыстории, предназначен для любой из N ОС;
- времена прохождения и обработки пакетов в локальных сетях вне радиосоты пренебрежимо малы;
- интенсивность помех, описываемая показателем BER (Bit Error Rate), постоянна и одинакова для всех станций.

Данные предположения обусловлены тем, что оценивается именно максимальная интенсивность TCP/IP-трафика, обеспечиваемая беспроводной сетью на основе протокола IEEE 802.11. В частности, эти предположения означают, что:

- освобожденное место в буфере БС мгновенно занимает следующий TCP/IP-пакетом, пункт назначения которого с равной вероятностью находится в любой из интегрированных локальных сетей;
- по получении TCP/IP-пакета от БС в очередь соответствующей ОС мгновенно помещается TCP-подтверждение на этот пакет.

В данном исследовании предполагается, что как БС, так и все ОС используют только механизм базового доступа ввиду нецелесообразности применения схемы RTS/CTS, поскольку вероятность разрушающей Б-коллизии очевидно мала, а TCP-подтверждения коротки.

7.4.1 Оценка пропускной способности

Для построения аналитической модели исследуемой беспроводной сети, состоящей из БС и N однородных ОС, введем ряд допущений в дополнение к тем, что были описаны в начале раздела.

Д.1. Кадры, передаваемые БС, одинаково воспринимаются всеми ОС.

Д.2. Время тайм-аута ОС является случайной величиной, распределенной по геометрическому закону с параметром $\nu = 1 - \sigma/\text{EIFS}$, т.е. оно равно $k > 0$ слотам отсрочки с вероятностью $(1 - \nu)\nu^{k-1}$.

Д.3. По завершении разрушающей О-коллизии все ее участники (и БС) блокированы, а вероятность неразрушающей О-коллизии пренебрежимо мала (ввиду малости SIFS и времени распространения сигнала δ).

Напомним, что станция блокирована, когда она ожидает истечения тайм-аута после неудачной передачи кадра DATA. Кроме того, будем говорить, что ОС пассивна, если ее очередь пуста, и активна в противном случае. Заметим, что БС не может отсчитывать свое время отсрочки и, следовательно, начинать передачу, пока хотя бы одна ОС блокирована. Теперь, как и в гл. 6, все время работы сети разобьем на неоднородные виртуальные слоты (с точки зрения БС, имеющих полное представление о происходящем в сети), каждый из которых может представлять собой:

- «пустой» слот длительностью σ — с момента окончания предыдущего слота ни одна из станций не начала передачу;
- «базовый успешный» слот (БУ-слот), в течение которого либо передает только БС, либо происходит неразрушающая Б-коллизия;
- слот Б-коллизии (БК-слот), в течение которого происходит разрушающая Б-коллизия; слот включает в себя последующий интервал EIFS;
- «оконечный успешный» слот (ОУ-слот), в течение которого только одна ОС ведет передачу;

- слот О-коллизии (ОК-слот), завершающийся переходом коллизирующих станций в блокированное состояние.

Обозначим вероятности этих слотов, очевидно зависящие от числа активных ОС n , через соответственно $p_{id}(n)$, $p_{bh}(n)$, $p_{bc}(n)$, $p_{eh}(n)$ и $p_{ec}(n)$. Как и в главе 6, название «успешный слот» дано условно, так как, несмотря на отсутствие разрушающей коллизии, попытка передачи может окончиться неудачей из-за искажения помехами. Вероятности искажения помехами кадра DATA с пакетом длиной r ($\xi_d(r)$) и кадра АСК (ξ_a) определяются исходя из интенсивности помех BER, т.е. кадр, состоящий из r байт, искажается с вероятностью $\xi_r = 1 - \exp\{-8r\text{BER}\}$. В свою очередь, эти вероятности определяют вероятность неудачной попытки передачи из-за искажения помехами для оконечной и базовой станций: $\xi_e = 1 - [1 - \xi_d(r_e)](1 - \xi_a)$ и $\xi_b = 1 - [1 - \xi_d(r_b)](1 - \xi_a)$, где r_e и r_b — длины TCP-подтверждения и TCP/IP-пакета (длина TCP/IP-пакета считается постоянной).

Поэтому с учетом возможных помех попытка БС успешна (и БУ-слот завершается интервалом DIFS) с вероятностью $P_{bh}(n) = p_{bh}(n)(1 - \xi_b)$ и неудачна (данный слот завершается интервалом EIFS) с вероятностью $1 - P_{bh}(n)$. Аналогично попытка ОС успешна (и ОУ-слот завершается интервалом DIFS) с вероятностью $P_{eh}(n) = p_{eh}(n)(1 - \xi_e)$, а с вероятностью $1 - P_{eh}(n)$ один из передаваемых кадров искажен коллизией или помехами, поэтому данный слот завершается переходом ОС в блокированное состояние. Заметим, что станции отслеживают дублирование принимаемой информации на MAC-уровне, и дубликаты ранее принятых пакетов, которые появляются при искажении кадров АСК, не передаются на IP-уровень. Легко показать, что успешно принятый пакет передается на IP-уровень с вероятностью $1 - \xi_a$.

Д.4. Пусть в очередях ОС ждут своей передачи ℓ TCP-подтверждений, причем $\ell = 0, \dots, L$, где $L = NL_B$, а L_B — размер буфера каждой ОС, измеряемый в числе TCP-подтверждений. Тогда все варианты размещения этих ℓ подтверждений по N очередям ОС сети равновероятны.

Опишем работу сети процессом рождения/гибели с дискретным временем, причем состояние этого процесса описывается значением ℓ , а единицей времени является виртуальный слот. Очевидно, интенсивности рождения и гибели определяются выражениями:

$$\lambda(0) = P_{bh}(0), \quad \lambda(\ell) = \sum_{n=1}^{\min(N, \ell)} P_{bh}(n) \gamma(n, \ell) [1 - \beta(n, \ell)],$$

$$\mu(\ell) = \sum_{n=1}^{\min(N, \ell)} P_{eh}(n) \gamma(n, \ell),$$

где $\gamma(n, \ell)$ и $\beta(n, \ell)$ — это соответственно вероятность наличия n активных ОС (при условии, что ℓ TCP-подтверждений ждут своей передачи) и вероятность отказа в передаче TCP-подтверждения оконечной станцией из-за переполнения ее буфера (при условии ℓ TCP-подтверждений, ожидающих своей передачи, и n активных ОС). Согласно допущению Д.4 эти вероятности рассчитываются сле-

дующим образом:

$$\gamma(n, \ell) = \frac{1}{g[\ell, N, L_B]} \binom{N}{k} g[\ell - n, n, L_B - 1],$$

$$\beta(n, \ell) = \frac{1}{g[\ell - n, n, L_B - 1]} \sum_{k=1}^{K(\ell, n)} \frac{k}{N} \binom{n}{k} g[\ell - n - k(L_B - 1), n - k, L_B - 2],$$

где $g[u, v, M]$ — число вариантов, которыми можно разместить u неразличимых шаров по v урнам так, что в любой урне было бы не более M шаров, а $K(\ell, n)$ — наибольшее целое, не превышающее отношения $(\ell - n)/(L_B - 1)$. Функция $g[u, v, M]$ вычисляется рекурсивно согласно [212]:

$$g[0, v, M] = 1 \quad \forall v > 0, \quad g[u, 1, M] = 1 \text{ при } u \leq M \text{ и } g[u, 1, M] = 0 \text{ при } u > M;$$

$$g[u, v, M] = \sum_{k=0}^{\min(u, M)} g[u - k, v - 1, M] \text{ при } v \geq 2, \quad u > 0.$$

Таким образом, средние количества пакетов, успешно переданных MAC-уровнем сети на IP-уровень за секунду в направлениях «сверху вниз» (S_{down}) и «снизу вверх» (S_{up}), находятся с помощью формул:

$$S_{down} = \frac{1}{T_{slot}} \left\{ P_{bh}(0)\pi(0) + \sum_{\ell=1}^{L-1} \pi(\ell) \sum_{n=1}^{\min(N, \ell)} P_{bh}(n)\gamma(n, \ell) \right\}, \quad (7.12)$$

$$S_{up} = \frac{1}{T_{slot}} \sum_{\ell=1}^L \mu(\ell)\pi(\ell), \quad (7.13)$$

где

$$T_{slot} = t_{slot}(0)\pi(0) + \sum_{\ell=1}^L \pi(\ell) \sum_{n=1}^{\min(N, \ell)} t_{slot}(n)\gamma(n, \ell) \quad (7.14)$$

— средняя длительность виртуального слота,

$$\pi(\ell) = G^{-1} \prod_{i=1}^{\ell} \frac{\lambda(i-1)}{\mu(i)}$$

— стационарные вероятности состояний, а G — нормирующая константа.

Теперь можно определить искомую пропускную способность сети S для ТСП-соединений, понимаемую как среднее число байт полезной информации ТСП/IP-уровня, переданных за секунду пользователям сети из внешнего мира:

$$S = r_b S_{up}. \quad (7.15)$$

Очевидно, при $L \ll 1$ $S_{down} \approx P_{bh}(N)$, $S_{up} \approx P_{ch}(N)$ и $T_{slot} \approx t_{slot}(N)$ в случае $P_{bh}(N) \geq P_{ch}(N)$, при этом $S_{down} > (1 - \xi_a) S_{up}$, т. е. передача значительной доли $\beta = 1 - S/(r_b S_{down})$ ТСП/IP-пакетов оказывается неподтвержденной из-за нехватки места в очередях ОС. При $P_{bh}(N) < P_{ch}(N)$ можно ограничиться вычислением небольшого числа слагаемых сумм (7.12) и (7.14) из-за их быстрой сходимости, причем в этом случае $S_{down} \approx S_{up}$ и можно считать, что передача всех ТСП/IP-пакетов подтверждается.

Итак, искомые пропускные способности могут быть вычислены при условии, что известны вероятности успешных слотов $P_{bh}(n)$ и $P_{eh}(n)$ и условная средняя длительность виртуального слота $t_{slot}(n)$ для всех $n = 0, \dots, N$. Для нахождения этих величин введем следующее допущение.

Д.5. Значения вероятностей БУ- и ОУ-слотов, а также условной средней длительности виртуального слота $t_{slot}(n)$, определяемые при условии n активных ОС, совпадают с соответствующими значениями, полученными для сети $\mathcal{S}_n$, аналогичной исследуемой, но включающей в себя только n ОС, к каждой из которых всегда имеются непустые очереди (по аналогии с [130] назовем такую сеть *насыщенной*).

7.4.2 Анализ насыщенной сети

Начнем с нахождения средней длительности виртуального слота $t_{slot}(n)$ для насыщенной сети $\mathcal{S}_n$. Очевидно,

$$t_{slot}(0) = \sigma p_{id}(0) + t_{bh}[1 - p_{id}(0)],$$

а при $n = 1, \dots, N$

$$t_{slot}(n) = \sigma p_{id}(n) + t_{bh}p_{bh}(n) + t_{bc}p_{bc}(n) + t_{eh}p_{eh}(n) + t_{ec}(n)p_{ec}(n), \quad (7.16)$$

причем длительность БК-слота (t_{bc}), а также средние длительности БУ- и ОУ-слотов (t_{bh} и t_{eh}) определяются выражениями

$$t_{bc} = t_{DATA}^b + \delta + \text{EIFS},$$

$$t_{bh} = t_{DATA}^b + \delta + [1 - \xi_d(r_b)](t_{ACK} + \text{SIFS} + \delta) + (1 - \xi_b)\text{DIFS} + \xi_b\text{EIFS},$$

$$t_{eh} = t_{DATA}^e + \delta + [1 - \xi_d(r_e)](t_{ACK} + \text{SIFS} + \delta) + (1 - \xi_e)\text{DIFS},$$

где t_{DATA}^b и t_{DATA}^e — времена передачи кадров DATA базовой и оконечными станциями.

Как и в [137], для оценки вероятностей слотов и средней длительности ОК-слота в насыщенной сети $\mathcal{S}_n$ будем использовать следующие допущения.

Д.6. Пусть в момент t станция в очередной раз уменьшает значение своего счетчика отсрочки. Тогда вероятность того, что значение счетчика станет равным нулю и, следовательно, станция начнет передачу, не зависит ни от предыстории, ни от поведения остальных станций и равна τ , если эта станция оконечная, и τ_b , если эта станция базовая.

Д.7. Вероятность коллизий, в которых участвуют более двух станций, пренебрежимо мала.

Наконец, для ограничения числа возможных состояний сети $\mathcal{S}_n$, описываемых числом заблокированных станций b , предположим следующее

Д.8. По завершении виртуального слота, не являющегося «пустым», все заблокированные станции, не пытавшиеся передавать в течение этого слота, покидают заблокированное состояние.

Из допущений Д.7 и Д.8 следует, что b может принимать только значения $0, \dots, B = \min\{2, n\}$, причем процесс изменения значений b описывается цепью Маркова с графом переходов, изображенным на рис. 7.14.

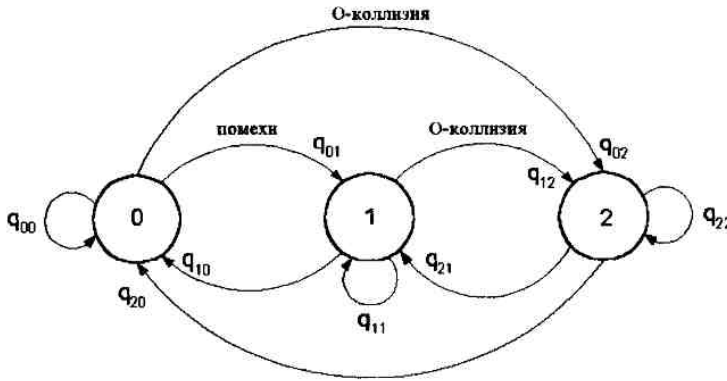


Рис. 7.14. Процесс изменения числа блокированных станций

Итак, найдем вероятности слотов в зависимости от числа блокированных станций b . Начнем со случая $b = 0$. Очевидно, вероятность пустого слота

$$\tilde{p}_{id}(n, 0) = (1 - \tau_b)(1 - \tau)^n,$$

а вероятности разрушающей и неразрушающей Б-коллизии и «базового успешного» слота равны соответственно:

$$\tilde{p}_{bc}(n, 0) = \tau_b \tau, \quad \tilde{p}_{bc}^0(n, 0) = \tau_b [1 - (1 - \tau)^n - \tau],$$

$$\tilde{p}_{bh}(n, 0) = \tau_b - \tilde{p}_{bc}(n, 0) = \tau_b(1 - \tau).$$

Следствием допущений Д.2 и Д.3 является синхронность ОС, означающая совпадение границ временных слотов неблокированных станций. Поэтому вероятность ОУ-слота равна:

$$\tilde{p}_{eh}(n, 0) = n\tau(1 - \tau_b)(1 - \tau)^{(v+1)(n-1)},$$

где v — наибольшее целое, не превышающее отношения t_{DATA}^e/σ . Кроме того, очевидно, вероятность разрушающей О-коллизии равна

$$\tilde{p}_{ec}(n, 0) = 1 - \tilde{p}_{id}(n, 0) - \tilde{p}_{bc}(n, 0) - \tilde{p}_{bh}(n, 0) - \tilde{p}_{eh}(n, 0)$$

$$(\tilde{p}_{ec}^0(n, 0) = \tilde{p}_{ec}(n, 0) = 0 \text{ при } n < 2).$$

Найдем также вероятность $\zeta(n, 0, i)$ ($i = 0, \dots, v$) того, что в начале данного виртуального слота стартует передача некоторой ОС, причем остальные станции воздержатся от передачи в течение, по крайней мере, $i\sigma$ при условии отсутствия передачи БС:

$$\zeta(n, 0, i) = n\tau(1 - \tau)^{(n-1)(i+1)}.$$

Состояния с $b > 0$ отличаются тем, что

1) в начале слота не могут передавать ни блокированные ОС, ни БС (поэтому $\tilde{p}_{bc}(n, b) = \tilde{p}_{bh}(n, b) = 0$ и, следовательно, вероятность пустого слота $\tilde{p}_{id}(n, b) = (1 - \tau)^{n-b}$);

2) в моменты $t = \sigma, 2\sigma, \dots, v\sigma$ с начала «непустого» слота каждая из блокированных станций может покинуть блокированное состояние.

Используя допущения Д.2 и Д.6, найдем вероятность η_i ($i = 0, \dots, v$) того, что в течение, по крайней мере, $i\sigma$ с начала передачи некоторой станции данная

блокированная станция не начнет свою передачу (приводя к коллизии):

$$\eta_i = \nu^i + \sum_{j=1}^i (1-\nu)\nu^{j-1}(1-\tau)^{i-j+1},$$

т.е. при $\nu \neq 1-\tau$

$$\eta_i = \nu^i + (1-\nu) \frac{(1-\tau)^i - \nu^i}{1 - \frac{\nu}{1-\tau}},$$

а при $\nu = 1-\tau$ $\eta_i = \nu^i[1 + i(1-\nu)]$.

Тогда вероятности ОУ- и ОК-слотов равны соответственно:

$$\tilde{p}_{eh}(n, b) = (n-b)\tau(1-\tau)^{(v+1)(n-b-1)}\eta_v^b, \quad \tilde{p}_{ec}(n, b) = 1 - \tilde{p}_{id}(n, b) - \tilde{p}_{eh}(n, 0).$$

Кроме того,

$$\zeta(n, b, i) = (n-b)\tau(1-\tau)^{(n-b-1)(i+1)}\eta_i^b.$$

Теперь, предполагая известными стационарные вероятности $\kappa(b)$ состояний $b = 0, \dots, B$ насыщенной сети $\mathcal{S}_n$, можно найти искомые вероятности слотов:

$$\begin{aligned} p_{id}(n) &= \sum_{b=0}^B \kappa(b) \tilde{p}_{id}(n, b), & p_{bh}(n) &= \sum_{b=0}^B \kappa(b) \tilde{p}_{bh}(n, b), \\ p_{bc}(n) &= \sum_{b=0}^B \kappa(b) \tilde{p}_{bc}(n, b), & p_{eh}(n) &= \sum_{b=0}^B \kappa(b) \tilde{p}_{eh}(n, b), \\ p_{ec}(n) &= \sum_{b=0}^B \kappa(b) \tilde{p}_{ec}(n, b), \end{aligned}$$

а также среднюю длительность слота О-коллизии:

$$t_{ec}(n) = t_{\text{DATA}}^e + \sigma \sum_{b=0}^B \frac{\kappa(b)}{1 - (1-\tau)^{n-b} - \zeta(n, b, v)} \sum_{i=0}^v i [\zeta(n, b, i-1) - \zeta(n, b, i)].$$

С учетом допущений Д.2, Д.3, Д.7 и Д.8 определим вероятности переходов q_{ij} , $i, j = 0, \dots, B$ в цепи Маркова на рис. 7.14. Сеть переходит из состояния $b = 0$ в состояние $b = 1$ при попытке ОС, окончившейся неудачей из-за помех, т.е. с вероятностью $q_{01} = \tilde{p}_{eh}(n, 0)\xi_e$, и в состояние $b = 2$ при О-коллизии (если $n > 1$), т.е. с вероятностью $q_{02} = \tilde{p}_{ec}(n, 0)$ (естественно, она остается в состоянии $b = 0$ с вероятностью $q_{00} = 1 - q_{01} - q_{02}$). Сеть остается в состоянии $b = 1$ в двух случаях: 1) «пустой» слот и блокированная ОС не становятся активными и 2) происходит попытка передачи ОС, оканчивающаяся неудачей из-за помех, т.е. $q_{11} = \tilde{p}_{id}(n, 1)\nu + \tilde{p}_{eh}(n, 1)\xi_e$. Сеть переходит из состояния $b = 1$ в состояние $b = 2$ при $n > 2$ в случае О-коллизии, т.е. $q_{12} = \tilde{p}_{ec}(n, 1)$ и $q_{10} = 1 - q_{11} - q_{12}$. Сеть остается в состоянии $b = 2$ в случае, если либо имеет место «пустой» слот и обе блокированные ОС не становятся активными, либо происходит О-коллизия, т.е. $q_{22} = \tilde{p}_{id}(n, 2)\nu^2 + \tilde{p}_{ec}(n, 2)$. Наконец, сеть переходит из состояния $b = 2$ в состояние $b = 0$, если либо обе блокированные ОС становятся активными в течение «пустого» слота, либо происходит удачная попытка передачи ОС, т.е. $q_{20} = \tilde{p}_{id}(n, 2)(1-\nu)^2 + \tilde{p}_{eh}(n, 2)(1-\xi_e)$ и $q_{21} = 1 - q_{22} - q_{20}$. Итак, решая систему

линейных уравнений

$$\kappa(b) = \sum_{i=0}^B q_{ib} \kappa(i), \quad b = 0, \dots, B-1, \quad \sum_{b=0}^B \kappa(b) = 1,$$

находим стационарные вероятности состояний $\kappa(b)$.

Осталось определить вероятности передачи τ и τ_b . С учетом допущения Д.6 имеем:

$$\tau = \frac{f_e}{f_e + \bar{w}_e}, \quad \tau_b = \frac{f_b}{f_b + \bar{w}_b},$$

где f_e и f_b — среднее число попыток, совершаемых соответственно оконечной и базовой станцией в процессе передачи одного пакета, а $\bar{w}_e$ и $\bar{w}_b$ — среднее число виртуальных слотов, в которые соответственно оконечная или базовая станция воздерживается от передачи в процессе передачи пакета. Предполагая неограниченным число попыток передачи любого пакета и учитывая правило установки времени отсрочки (6.2), находим эти средние числа для ОС:

$$f_e = \frac{1}{1 - \chi_e}, \quad (7.17)$$

$$\begin{aligned} \bar{w}_e &= (1 - \chi_e) \sum_{i=1}^{m+1} \chi_e^{i-1} \sum_{j=0}^{i-1} \frac{W_0 2^j - 1}{2} + \\ &+ (1 - \chi_e) \sum_{i=m+2}^{\infty} \chi_e^{i-1} \left\{ \sum_{j=0}^m \frac{W_0 2^j - 1}{2} + (i - m - 1) \frac{W_0 2^m - 1}{2} \right\} = \\ &= (1 - \chi_e) \sum_{i=1}^{m+1} \chi_e^{i-1} \left(W_{i-1} - \frac{W_0 + i}{2} \right) + \\ &+ \frac{\chi_e^{m+1}}{2} \left\{ (W_m - 1) \frac{m + 2 - (m + 1)\chi_e}{1 - \chi_e} - (m - 1)W_m - W_0 \right\}. \end{aligned} \quad (7.18)$$

Видно, что эти средние числа зависят от вероятности χ_e того, что попытка, совершаемая ОС, окажется неудачной. Аналогичные формулы (с заменой вероятности χ_e на соответствующую вероятность χ_b) справедливы и для БС, т. е. для значений f_b и $\bar{w}_b$. Относя среднее число успешных попыток к общему числу попыток, совершаемых соответственно оконечной и базовой станцией в течение одного виртуального слота (с учетом допущения Д.7), находим вероятности χ_e и χ_b :

$$\chi_e = 1 - \frac{(1 - \xi_e) \sum_{b=0}^B \kappa(b) \tilde{p}_{eh}(n, b)}{\sum_{b=0}^B \kappa(b) [\tilde{p}_{eh}(n, b) + 2\tilde{p}_{ec}(n, b) + \tilde{p}_{bc}(n, b) + \tilde{p}_{bc}^0(n, b)]}, \quad (7.19)$$

$$\chi_b = 1 - \frac{(1 - \xi_b) \tilde{p}_{bh}(n, 0)}{\tau_b} = 1 - (1 - \xi_b)(1 - \tau). \quad (7.20)$$

Таким образом, вероятности передачи τ_b и τ для каждой из насыщенных сетей S_n оцениваются с помощью следующей итеративной технологии.

- Шаг 0. Задать некоторые исходные значения τ_b и τ .
- Шаг 1. Для каждого числа $b = 0, 1, 2$ блокированных ОС вычислить вероятности всех возможных типов виртуальных слотов: $\tilde{p}_{id}(n, b)$, $\tilde{p}_{bh}(n, b)$, $\tilde{p}_{eh}(n, b)$, $\tilde{p}_{ec}(n, b)$, $\tilde{p}_{bc}(n, b)$, $\tilde{p}_{bc}^0(n, b)$.
- Шаг 2. Вычислить вероятности переходов q_{ij} , $i, j = 0, 1, 2$, и стационарные вероятности $\pi(b)$, $b = 0, 1, 2$.
- Шаг 3. Вычислить сначала вероятности неудачи попытки (χ_b и χ_e) по формулам (7.19) и (7.20), затем по формулам (7.17) и (7.18) — количество попыток (f_b и f_e) и виртуальных слотов ($\tilde{w}_b$ и $\tilde{w}_e$), в которые БС и ОС воздерживаются от передачи.
- Шаг 4. Найти модифицированные значения τ_b и τ и сравнить эти значения с исходными. В случае, если различие этих значений превышает заранее установленный предел, возвратиться к шагу 1, приняв в качестве новых исходных значений τ_b и τ полусуммы их старых исходных и модифицированных значений.

После выполнения этой итерационной процедуры находятся усредненные вероятности каждого типа слотов и средняя длительность виртуального слота (по формулам (7.14) и (7.16)), а затем по формулам (7.12) и (7.15) — искомые пропускные способности.

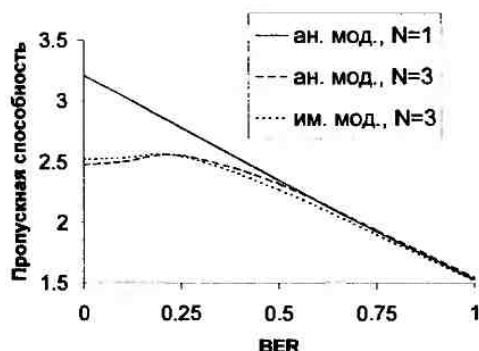
7.4.3 Численные результаты

В этом подразделе разработанные аналитические методы (совместно с имитационным моделированием на базе языка GPSS [157, 195]) используются для численного анализа зависимости пропускной способности S от интенсивности помех BER и числа ОС N . Объектом исследования является региональная сеть, станции которой взаимодействуют под управлением высокоскоростной версии 802.11b (короткая преамбула) [161] протокола IEEE 802.11. Значения параметров протокола и трафика, используемые при моделировании, приведены в табл. 7.4. Они были определены на основании [161, 222], а также (в том, что касается размеров TCP/IP-пакетов и подтверждений) [57, 80].

Рассмотрим сначала графики на рис. 7.15. Видно, что пропускная способность быстро снижается с ростом BER в случае единственной ОС. При нескольких ОС изменение пропускной способности немонотонно: при отсутствии помех среднее время, затрачиваемое на передачу TCP-подтверждения, оказывается больше, чем время передачи TCP/IP-пакета, из-за коллизий, что вызывает появление очередей на ОС. Таким образом, небольшой рост BER оказывается даже «полезен» в смысле повышения пропускной способности, так как в этом случае ОС получают достаточно времени, чтобы успеть послать TCP-подтверждение до получения следующего TCP/IP-пакета, и очереди не образуются. Этот достаточно неожиданный вывод подтверждается имитационным моделированием, которое также показало хорошую точность разработанного аналитического метода (см. графики для $N = 3$ на рис. 7.15).

Таблица 7.4. Параметры протокола и трафика

Параметр	Значение
Скорость канала, V	11 Мбит/с
Минимальное окно, W_0	32
Максимальное окно, W_m	1024
Слот, σ	20 мкс
SIFS	10 мкс
DIFS	50 мкс
EIFS	212 мкс
Размер ACK	29 байт
t_{ACK}	106 мкс
PHY+MAC-заголовок, H	49 байт
TCP/IP-данные, r_b	576 байт
t_{DATA}^b	540 мкс
TCP-подтверждения, r_e	80 байт
t_{DATA}^e	179 мкс

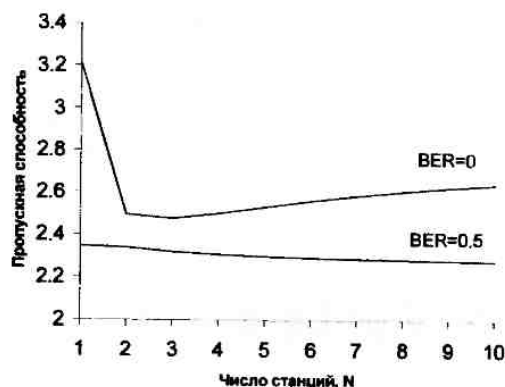
Рис. 7.15. Зависимости пропускной способности S (Мбит/с) от $BER \times 10^4$, полученные аналитическим (ан. мод.) и имитационным (им. мод.) моделированием

Теперь рассмотрим изменение пропускной способности при увеличении числа ОС N (рис. 7.16). При высоком BER ($BER = 5 \times 10^{-5}$ на рис. 7.16) пропускная способность уменьшается лишь слегка, оставаясь примерно на одном и том же уровне, так как время передачи TCP/IP-пакета велико и ОС имеют достаточно времени на отправку TCP-подтверждения до получения следующего TCP/IP-пакета. Поэтому в сети обычно только одна ОС активна и коллизии не возникают.

При отсутствии помех картина совершенно меняется. Очевидно, пропускная способность максимальна при единственной ОС ввиду отсутствия коллизий. При среднем числе ОС пропускная способность довольно мала ввиду высокой вероятности коллизий. В этом случае ОС работают фактически в режиме насыщения и не могут обеспечить скорость передачи пакетов «снизу вверх» S_{up} , адекватную скорости передачи пакетов «сверху вниз» S_{down} , что приводит к частым потерям TCP-подтверждений из-за переполнения очередей на ОС. Наконец, при

дальнейшем увеличении числа ОС N они выходят из режима насыщения, скорости «снизу вверх» и «сверху вниз» становятся примерно равными, и пропускная способность медленно растет.

Рис. 7.16. Зависимости пропускной способности S (Мбит/с) от числа ОС N при разных значениях $BER \times 10^4$



7.5 Применение централизованного управления

Как показывают результаты, приведенные в предыдущем разделе, мощность беспроводного канала в случае нескольких скрытых ОС используется весьма неэффективно: при скорости канала, равной 11 Мбит/с, пропускная способность, обеспечиваемая для интернет-соединений, не превышает 3,2 Мбит/с даже при отсутствии коллизий, а при их наличии — 2,5 Мбит/с. Причиной этого являются существенные потери производительности, неизбежно присущие децентрализованной схеме DCF и связанные с большим временем отсрочки и высокой вероятностью коллизий.

Поэтому применительно к исследуемой архитектуре региональной сети (см. рис. 7.1) естественно оценить эффективность ее работы при централизованном управлении, обеспечиваемом функцией централизованного управления (Point Coordination Function — PCF). При механизме PCF, предусмотренном стандартом IEEE 802.11 в качестве альтернативы DCF, станция-координатор (в нашем случае — БС) ведет централизованный опрос остальных станций (см. рис. 7.17). На БС хранится список, в соответствии с которым она выбирает следующую по очереди ОС и посылает ей кадр, содержащий приглашение к передаче, а также пакет данных, предназначенный для этой станции, если он имеется. Кроме того, если на предыдущем шаге опроса БС приняла пакет данных, то в данный кадр включается (путем изменения соответствующего бита заголовка) также подтверждение на этот пакет.

ОС, получившая приглашение к передаче, передает спустя короткий интервал SIFS ответный кадр, который содержит пакет данных, если он имеется (при пустой очереди ОС она отвечает так называемым нулевым кадром NULL, состоящим только из заголовка), а также подтверждение на прием пакета от БС

(если он был). По получении этого кадра БС спустя SIFS переходит к опросу следующей по списку ОС. В случае неудачной передачи опросного кадра и, как следствие, отсутствия ответного кадра в течение тайм-аута PIFS начинается следующий шаг опроса.

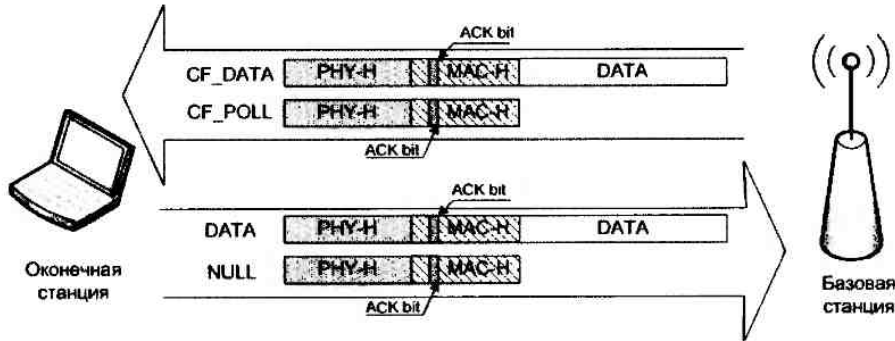


Рис. 7.17. Типы кадров протокола 802.11 PCF

Так как все ОС считаются однородными, предположим, что они опрашиваются строго по кругу. Разделим все время работы сети на PCF-слоты, каждый из которых начинается и заканчивается в моменты начала опроса ОС, соседних в списке. В течение PCF-слота с вероятностью $1 - \xi_d(r_b)$ происходит успешная передача ТСР-пакета к ОС j , который после проверки на дублирование (доля дубликатов, очевидно, равна $\xi_d(r_e)$) передается на IP-уровень, в результате чего (в соответствии с допущением о мгновенном обслуживании) в очередь ОС поступает ТСР-подтверждение. Таким образом, вероятность α_b поступления нового ТСР-подтверждения в течение PCF-слота равна $\alpha_b^0 = [1 - \xi_d(r_b)][1 - \xi_d(r_e)]$. Аналогично с вероятностью α_e ТСР-подтверждение считается успешно переданным и покидает очередь ОС в течение PCF-слота, причем для стандартной схемы PCF $\alpha_e = \alpha_e^0 = [1 - \xi_d(r_b)]^2 [1 - \xi_d(r_e)]$. Так как $\alpha_b^0 \geq \alpha_e^0$ и размер буфера $L \gg 1$, очереди ОС можно считать насыщенными. Следовательно, искомая пропускная способность определяется выражением:

$$S = r_b \alpha_e / t_{slot}^{PCF}, \quad (7.21)$$

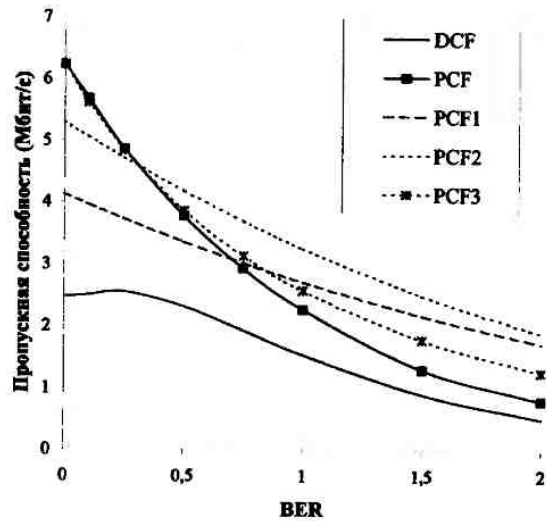
где t_{slot}^{PCF} — средняя длительность PCF-слота, причем для рассматриваемой стандартной схемы PCF $\alpha_e = \alpha_e^0$ и

$$t_{slot}^{PCF} = t_{slot}^{PCF0} = t_{DATA}^b + \xi_d(r_b) PIFS + [1 - \xi_d(r_b)] [t_{DATA}^e + 2(\delta + SIFS)]. \quad (7.22)$$

На рис. 7.18 приведены графики сравнения пропускной способности, обеспечиваемой при распределенном (кривая DCF) и централизованном (кривая PCF) управлении в случае $N = 3$ ОС. Видно, что при слабых помехах PCF превосходит по производительности DCF более чем в 2 раза. Еще большего эффекта при слабых помехах можно добиться, объединяя последовательные пакеты (и ответные ТСР-подтверждения), предназначенные для одной и той же ОС, в один MAC-кадр: на рис. 7.19 кривые для разных вариантов PCF получены при объединении пакетов, оптимизированном для конкретного BER. Это объединение

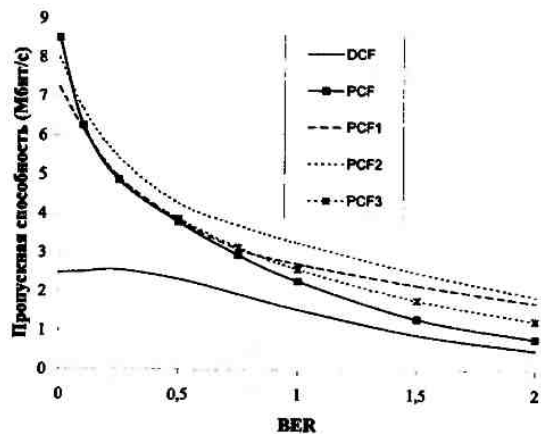
пакетов значительно легче осуществить в PCF, так как при этой схеме очередь пакетов в БС не едина, а разбита на очереди по направлениям.

Рис. 7.18. Сравнение пропускной способности DCF и PCF в зависимости от $BER \times 10^4$ при $N = 3$



При сильных помехах разница в пропускной способности, обеспечиваемой DCF, и основного варианта PCF (кривые PCF на рис. 7.18 и 7.19) не так велика, что обусловлено большой вероятностью неудачного обмена. В этом случае «прикрепление» приглашения к передаче и MAC-подтверждения к кадрам данных может оказаться нецелесообразным. Поэтому рассмотрим следующие альтернативные PCF-схемы, также не противоречащие стандарту IEEE 802.11.

Рис. 7.19. Сравнение пропускной способности DCF и PCF с объединением пакетов



- При схеме PCF1 приглашения к передаче и MAC-подтверждения передаются отдельными кадрами, т. е. в общем случае каждый PCF-слот содержит до пяти фаз:

1. БС передает кадр DATA к ОС.
2. ОС, успешно принявшая этот кадр, отвечает кадром ACK.
3. БС посылает опросный кадр CF-POLL без данных.
4. ОС отвечает (спустя SIFS) кадром DATA (или NULL).
5. БС посылает подтверждающий кадр CF-ACK.

Все эти фазы разделены интервалами SIFS. БС ожидает ответа от ОС после фаз 1 и 3 в течение PIFS.

- **Схема PCF2** использует отдельный кадр CF-ACK для подтверждения успешной передачи данных от ОС.
- При **схеме PCF3** этот кадр CF-ACK используется только для подтверждения принятых (от ОС) дубликатов.

При схемах PCF2 и PCF3 $\alpha_b = \alpha_b^0$, а вероятность α_e равна соответственно $\alpha_e^2 = [1 - \xi_d(r_b)][1 - \xi_d(r_e)](1 - \xi_0)$ или $\alpha_e^3 = [1 - \xi_d(r_b)][1 - \xi_d(r_e)](1 - D)$,

где ξ_0 — вероятность искажения кадра CF-ACK (а также кадров CF-POLL и NULL, имеющих тот же размер, равный H), а $D = \xi_d(r_b)/[1 - \xi_0 + \xi_d(r_b)]$ — доля дубликатов в общем числе TCP-подтверждений, принятых БС, для схемы PCF3. Очевидно, при этих схемах очереди ОС также являются насыщенными, поэтому пропускная способность определяется (7.21) с $\alpha_e = \alpha_e^i$ и $t_{slot}^{PCFi} = t_{slot}^{PCFi}$ для схемы PCFi, $i = 2, 3$, причем t_{slot}^{PCFi} вычисляется по формуле (7.22) с добавлением в последнюю квадратную скобку слагаемого $[1 - \xi_d(r_e)](t_0 + \text{SIFS} + \delta)$ для PCF2 или $[1 - \xi_d(r_e)]D(t_0 + \text{SIFS} + \delta)$ для PCF3, где t_0 — время передачи кадра CF-ACK (а также кадров CF-POLL и NULL), равное 121 мкс при скорости $V = 11$ Мбит/с.

При схеме PCF1

$$\alpha_b = \alpha_b^1 = [1 - \xi_d(r_b)](1 - \xi_a) \quad \text{и} \quad \alpha_e = \alpha_e^1 = (1 - \xi_0)^2[1 - \xi_d(r_e)],$$

причем при $\text{BER} > 0$ $\alpha_b^1 < \alpha_e^1$ (так как $\xi_d(r_b) \gg \xi_d(r_e)$) и очереди ОС не насыщены. Очевидно, за цикл очередь TCP-подтверждений произвольно выбранной ОС увеличивается на 1 с вероятностью $\lambda = \alpha_b^1(1 - \alpha_e^1)$ и уменьшается на 1 с вероятностью $\mu = \alpha_e^1(1 - \alpha_b^1)$. Рассматривая изменение размера очереди ℓ как процесс рождения/гибели, находим вероятность $\rho = \lambda/\mu$ непустой очереди в момент начала PCF-слота, относящегося к данной ОС. В результате (7.21) преобразуется к виду:

$$S = r_b \rho_e \alpha_e^1 / t_{slot}^{PCF1},$$

где $\rho_e = (1 - \rho)\alpha_b^1 + \rho$ — вероятность непустой очереди в момент ее опроса, а

$$t_{slot}^{PCF1} = t_{\text{DATA}}^b + \xi_d(r_b)\text{PIFS} + [1 - \xi_d(r_b)][t_{\text{ACK}} + 2(\delta + \text{SIFS})] + t_0 + \xi_0\text{PIFS} + (1 - \xi_0)\{\rho_e t_{\text{DATA}}^e + (1 - \rho_e)t_0 + 2(\delta + \text{SIFS}) + \rho_e[1 - \xi_d(r_e)](t_0 + \delta + \text{SIFS})\}.$$

При отсутствии помех ($\text{BER} = 0$) эти формулы остаются верны с $\rho_e = 1$.

Зависимости пропускной способности S от BER для альтернативных схем PCFi, полученные с помощью формул этого раздела, также приведены на рис. 7.18

и 7.19. Видно, что при объединении пакетов схему PCF2 можно считать оптимальной (стандартная схема PCF выигрывает у схемы PCF2 меньше 5% и только при отсутствии помех). Если же пакеты передаются без объединения, стандартная схема PCF оптимальна при слабых помехах ($BER < 3 \cdot 10^{-5}$), а при более сильных помехах оптимальной является схема PCF2 с передачей подтверждения отдельным кадром. В связи с трудностью определения BER переключение со стандартной схемы PCF на схему PCF2 можно проводить при достижении определенного процентного соотношения дубликатов пакетов, принимаемых от ОС. В соответствии с рис. 7.18 и принятым размером TCP-пакета это предельное соотношение составляет 14%, т. е. на пакет приходится в среднем 0,14 дубликата при скорости канала $V = 11$ Мбит/с.

7.6 Динамический опрос в беспроводных сетях с централизованным управлением

Эффективность схемы DCF высока при низкой загруженности сети, однако производительность этой схемы существенно падает при увеличении количества терминалов и нагрузки, что связано с высокой вероятностью коллизий и увеличенными временами отсрочки. Кроме того, в типичных условиях региональной беспроводной сети оконечные станции не имеют радиовидимости друг с другом (т. е. скрыты друг от друга) и вынуждены взаимодействовать через ретрансляционную базовую станцию, что дополнительно увеличивает вероятность коллизий.

Механизм PCF позволяет решить эти проблемы, так как он работает в условиях отсутствия конкуренции, что позволяет обеспечивать гораздо более высокий максимум пропускной способности, чем DCF [143, 167, 189, 210]. К сожалению, в настоящее время радиоустройства, реализующие механизм PCF, не производятся, по крайней мере серийно. Это связано с тем, что ранее протокол IEEE 802.11 использовался при создании только локальных сетей, где отсутствуют скрытые станции и производительность DCF значительно выше. В связи с этим в Институте проблем передачи информации РАН начата разработка нового радиосредства, реализующего схему PCF. Описание этого радиосредства см. в разделе 9.2.

7.6.1 Стратегии опроса

В общем виде PCF описывается следующей схемой (см. рис. 7.17): базовая станция, называемая также точкой доступа (ТД), опрашивает терминалы (оконечные станции) по кругу, т. е. PCF представляет собой схему TDMA, в которой время работы сети разбито на циклы опроса, состоящие из временных слотов переменной длины (см. рис. 7.20). Слот i предназначен для обмена кадрами между ТД и i -м терминалом. В отличие от терминала, контролирующего одну очередь пакетов, ТД управляет N очередями ТД (ОТД), где N — количество опрашиваемых терминалов, и ОТД j содержит пакеты, предназначенные для передачи терминалу j . Назовем ОТД j и очередь j -го терминала (ОТ j) противоположными очередями и предположим, что их размеры не ограничены.

При статическом методе опроса, описанном в начале предыдущего раздела и изученном в [3], в случае большого количества терминалов и невысоком или пульсирующем трафике происходят значительные потери пропускной способности, вызванные неудачными попытками опроса, в ответ на которые не последовала передача полезной информации [143, 167, 189]. В этом весьма распространенном случае механизм PCF оказывается менее эффективен, чем DCF, и именно поэтому он не нашел широкого применения до настоящего времени. Данный раздел посвящен методам уменьшения указанных потерь. Предшествующая попытка решения этой проблемы была предпринята в [189], где была предложена схема скрытой сигнализации, в соответствии с которой терминал обозначает («взводит» бит, специально добавленный в MAC-заголовок), что его очередь не пуста. Однако этот подход, во-первых, приводит к потере совместимости со стандартными устройствами, поддерживающими протокол IEEE 802.11, и, во-вторых, полагается на схему DCF при решении проблемы возобновления опросов терминала. Предлагаемый нами подход к решению этой проблемы использует только средства PCF. Более конкретно, предлагается использовать, развить и изучить стратегию отложенного опроса, предложенную для сетей Bluetooth в [135].

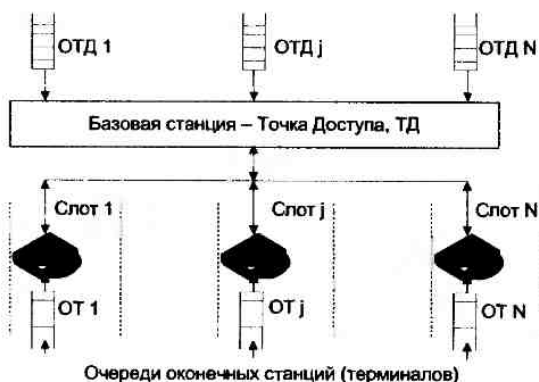


Рис. 7.20. Схема работы PCF

Если соответствующая ОТД не пуста (в противном случае слоты, предназначенные для этого терминала, будут пустыми, и ТД каждый цикл будет увеличивать значение счетчика k на 1). При получении от терминала пакета с данными ТД возвращает его на нулевую стадию. Когда $k = W_i$, ТД обязательно опрашивает терминал и в случае пустого отклика увеличивает i на 1 (если $i < I$), устанавливая $k = 1$. Частный случай этой стратегии опроса (с $W_i = 2^i$) был предложен в [135] для сетей Bluetooth.

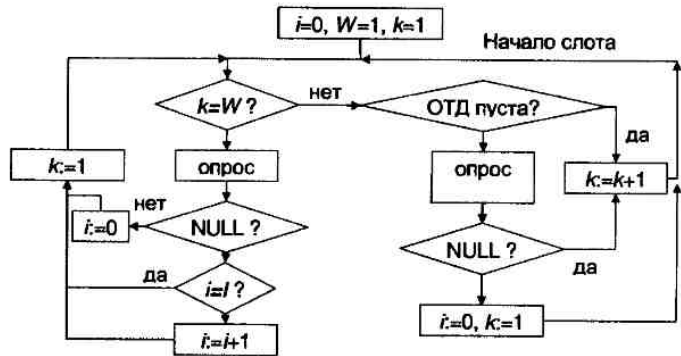
7.6.2 Аналитическая модель динамической стратегии отложенного опроса

В этом подразделе мы построим модели Маркова, описывающие изменения очередей в сети 802.11 PCF в случае идеального канала и обобщенной стратегии

В соответствии с этой динамической стратегией (см. рис. 7.21) терминал непременно опрашивается, только если его счетчик отсрочки k равен окну опроса W_i , заданному для каждой стадии его опроса $i = 0, 1, \dots, I$. В нулевой стадии $k = W_0 = 1$, и терминал опрашивается в каждом цикле, находясь в Λ_2 . Когда ТД получает нулевой кадр от терминала, она определяет, что ОТ пуста, и устанавливает $i = 1$. В течение следующих $W_i - 1$ циклов ($1 < W_i \leq W_{i+1}$ для всех $0 < i < I - 1$) ТД будет опрашивать терминал, только ес-

динамического опроса. В целях учета пульсирующего характера поступающего трафика нагрузка на каждую очередь сети описана групповым пуассоновским потоком пакетов, причем количество пакетов в каждой группе распределено геометрически. С помощью этих моделей оценим средние времена обслуживания и пребывания пакета в каждой очереди, которые являются основными показателями производительности в режиме нормальной нагрузки. В частности, среднее время обслуживания определяется как среднее время между либо приемами подтверждений на последовательные пакеты в очереди (если пакет приходит в непустую очередь), либо моментами прибытия пакета и подтверждения его передачи. Оба эти показателя оказывают существенное влияние на производительность транспортных протоколов, таких как TCP.

Рис. 7.21. Обобщенная динамическая стратегия отложенного опроса



Объект исследования — сеть IEEE 802.11 PCF, состоящая из ТД и N терминалов. Интенсивность прибытия групп пакетов, среднее время передачи пакета (включая MAC- и PHY-заголовки) и средний размер группы равны соответственно Λ_d , T_d и $m_d = q_d^{-1}$ для ОТД и Λ_u , T_u и $m_u = q_u^{-1}$ для ОТ. Пусть $l_{dj}(t)$ и $v_j(t) = [i_j(t), k_j(t), l_{uj}(t)]$ — случайные процессы, отображающие состояния ОТД j и ОТ j в момент времени t . Состояние ОТД описывается только длиной ОТД l_{dj} , измеренной в группах, в то время как описание состояния ОТ (при адаптивном опросе) включает также номер стадии опроса $i_j = 0 \dots I$ и значение счетчика опросов $k_j = 1 \dots W_i$. При статическом опросе, состояние ОТ описывается только с помощью l_{uj} . Пусть $\pi_d(l)$ и $\pi_u(i, k, l)$ — стационарные вероятности этих состояний.

Для обоих процессов примем дискретную временную шкалу, единицей измерения которой является цикл. Для $l_{dj}(t)$ каждый момент t соответствует началу слота, предназначенного для j -го терминала, включая случай нулевого слота. Для $v_j(t)$ каждый момент t соответствует либо окончанию опроса терминала (если слот ненулевой), либо окончанию предыдущего слота. Предположим, что все $l_{dj}(t)$ и $v_j(t)$ независимы. Однако на самом деле $v_j(t)$ зависит от $l_{dj}(t)$, так как условно опрашиваемый терминал (т. е. при $k < W_i$) может передавать, только если противоположная ОТД не пуста. Эта зависимость будет учтена путем подбора соответствующих переходных вероятностей для $v_j(t)$. При моделировании примем следующее.

Основное допущение: для любой очереди вероятность поступления более чем одной группы пакетов в течение цикла пренебрежимо мала.

Модель очереди точки доступа. Очевидно, что $l_{dj}(t)$ — процесс рождения/гибели, где «рождение» происходит, когда за цикл поступает новая группа, а обслуживание текущей группы еще не закончено. Таким образом, вероятность «рождения» равна $\lambda_d^0 = 1 - \exp\{-\Lambda_d(T_c^* + T_{sl}^d(0))\}$ (для $l_d = 0$), а при $l_d > 0$

$$\lambda_d(l_d) = (1 - q_d)[1 - \exp\{-\Lambda_d(T_c^* + T_{sl}^d(l_d))\}], \quad (7.23)$$

где $T_{sl}^d(l_d)$ — средняя продолжительность j -го слота, зависящая от l_d , а T_c^* — средняя продолжительность остальных $N - 1$ слотов. При $l_d = 0$ j -й слот является непустым при условии A , что терминал обязательно опрашивается в данном цикле, поэтому $T_{sl}^d(0) = T_{sl}^{d0} = \nu_p[t_0 + 2\delta + (1 - \rho_u^p)t_0 + \rho_u^p T_u]$, где t_0 — время передачи кадров CF-POLL или NULL, $\delta = \text{SIFS}$, ρ_u^p — вероятность того, что противоположная ОТ пуста при условии A , и ν_p — вероятность выполнения этого условия, т. е.:

$$\nu_p = 1 - \sum_{i=1}^I \sum_{k=1}^{W_i-1} \sum_{l=0}^k \pi_u(i, k, l), \quad \rho_u^p = \left[\nu_p - \sum_{i=0}^I \pi_u(i, W_i, 0) \right] / \nu_p.$$

При статическом опросе $\nu_p = 1$.

При $l_d > 1$ ОТ j не может находиться в таких состояниях

$$v_j(t) = [i_j(t) > 0, k_j(t), l_{uj}(t) > 0],$$

что $l_{uj}(t-1) > 0$ (пусть X — множество таких состояний $v_j(t)$), так как в противном случае этот терминал был бы опрошен в предыдущем цикле. Поэтому

$$T_{sl}^d(l_d > 1) = T_{sl}^{d1} = T_d + 2\delta + (1 - \rho_u^1)t_0 + \rho_u^1 T_u, \quad (7.24)$$

где ρ_u^1 — вероятность непустой ОТ j при условии $v_j \notin X$. Наконец, при $l_{dj}(t)=1$ длительность j -го слота зависит от $l_{dj}(t-1)$: если $l_{dj}(t-1) = 0$, то состояние ОТ j может быть любым, а при $l_{dj}(t-1) > 0$ должно выполняться $v_j \notin X$. Поэтому $T_{sl}^d(1) = T_{sl}^{d*}$ определяется той же формулой (7.24) с заменой ρ_u^1 на ρ_u^* , причем значение ρ_u^* будет определено ниже. Таким образом, $\lambda_d(1) = \lambda_d^*$ и $\lambda_d(l_d > 1) = \lambda_d$, где правые части этих равенств определяются подстановкой в (7.23) значений T_{sl}^{d*} и T_{sl}^{d1} .

«Гибель» происходит при завершении обслуживания текущей группы и неприбытии новой группы в данном цикле, так что ее вероятность равна

$$\mu_d^* = q_d \exp\{-\Lambda_d(T_c^* + T_{sl}^{d*})\} \quad \text{при } l_d = 1,$$

$$\mu_d = q_d \exp\{-\Lambda_d(T_c^* + T_{sl}^{d1})\} \quad \text{при } l_d > 1.$$

Наконец,

$$T_c^* = (N - 1)\{\pi_d(0)T_{sl}^{d0} + \pi_d(1)T_{sl}^{d*} + [1 - \pi_d(0) - \pi_d(1)]T_{sl}^{d1}\}.$$

Таким образом, находим стационарные вероятности:

$$\pi_d(0) = 1 - \rho_d = G_d^{-1}, \quad \pi_d(1) = G_d^{-1} \frac{\lambda_d^0}{\mu_d^*}, \quad (7.25)$$

$$\pi_d(l > 1) = G_d^{-1} \frac{\lambda_d^0 \lambda_d^*}{\mu_d^* \mu_d} \left(\frac{\lambda_d}{\mu_d} \right)^{l-2}, \quad (7.26)$$

где нормализующий множитель

$$G_d = 1 + \frac{\lambda_d^0}{\mu_d^*} \left[1 + \frac{\lambda_d^*}{\mu_d - \lambda_d} \right], \quad (7.27)$$

а ρ_d — вероятность непустой ОТД. Очевидно, должно выполняться условие $\lambda_d < \mu_d$.

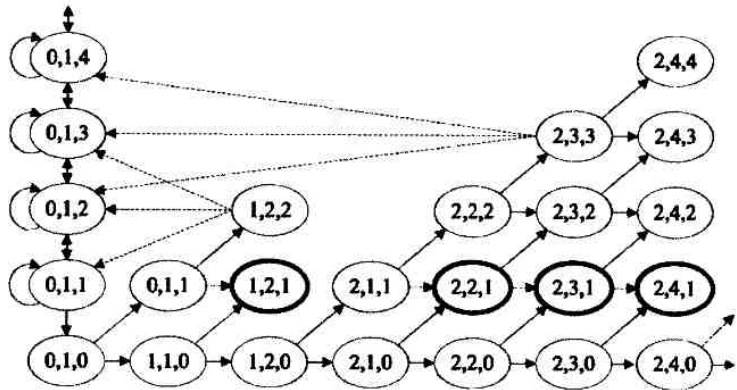
Теперь можно определить вероятность ρ_u^* . Так как вероятность того, что данная ОТД, длина которой равна 1, была пуста в предыдущем цикле, равна

$$\lambda_d^0 \pi_d(0) / \{ \lambda_d^0 \pi_d(0) + [1 - \lambda_d^* - \mu_d^*] \pi_d(1) + \mu_d \pi_d(2) \} = \mu_d^*,$$

то $\rho_u^* = \mu_d^* \rho_u + (1 - \mu_d^*) \rho_u^1$, где $\rho_u = 1 - \sum_{i=0}^I \sum_{k=1}^{W_i} \pi_u(i, k, 0)$ — безусловная вероятность того, что противоположная ОТ пуста. Значение ρ_u^1 определим далее при анализе модели ОТ.

Модель очереди терминала. При статическом опросе $l_{uj}(t)$ — также процесс рождения/гибели, стационарные вероятности которого определяются формулами (7.25)–(7.27), в которых используются λ^0 вместо λ_d^0 , λ вместо λ_d и λ_d^* и μ вместо μ_d и μ_d^* , которые, в свою очередь, определяются теми же формулами при подстановке в них Λ_u и q_u вместо Λ_d и q_d , $T_{sl}^{u1} = T_u + \delta + t^p(\rho_d)$ вместо T_{sl}^{d*} и T_{sl}^{d1} и $T_{sl}^{u0} = t_0 + \delta + t^p(\rho_d)$ вместо T_{sl}^{d0} [где $t^p(\rho_d) = (1 - \rho_d)t_0 + \rho_d T_d + \delta$ — среднее время опроса]. В дальнейшем рассматривается только динамический опрос.

Рис. 7.22. Начало цепи Маркова для ОТ с $W_1 = 2$ и $W_2 = 4$



Учитывая, что $l_u \leq k$ при $i > 0$, процесс $v_j(t)$ может рассматриваться как цепь Маркова, пример которой показан на рис. 7.22 (переходы, возвращающие терминал на нулевую стадию, показаны только для состояний (1,2,2) и (2,3,3)). Определим ненулевые вероятности одношаговых переходов, каждый из которых можно отнести к одному из следующих видов.

1. Увеличение счетчика пропущенных циклов (для $k < W_i$) или переход к следующей стадии (для $k = W_i$ и $l = 0$) без (γ -переходы) и с (ψ -переходы) увеличением длины ОТ. Вероятности этих переходов равны соответственно

$$\Gamma(\theta, \eta_1) = (1 - \theta) \exp\{-\Lambda_u(T_c^* + \eta_1)\} \text{ и } \Psi(\theta, \eta_1) = 1 - \theta - \Gamma(\theta, \eta_1)$$

при $l > 0$ и

$$\Gamma_0(\theta, \eta_0) = \exp\{-\Lambda_u[T_c^* + \eta_0 + \theta(t_0 + \delta)]\} \text{ и } \Psi_0(\theta, \eta_0) = 1 - \Gamma_0(\theta, \eta_0)$$

при $l = 0$, где θ и η (с различными индексами) — вероятность опроса и условное среднее значение времени следующего опроса. Заметим, что после γ - и ψ -переходов из $(I, W_I, 0)$ ОТ оказывается соответственно в $(I, 1, 0)$ и $(I, 1, 1)$.

2. Для состояний с $l > 0$: переход в состояние нулевой стадии с увеличением, уменьшением и без изменения длины ОТ (α -, ϕ - и β -переходы). Их вероятности равны:

$$A(\theta, \eta_2) = \theta(1 - q_u)[1 - \exp\{-\Lambda_u(T_c^* + \eta_2 + T_u + \delta)\}],$$

$$\Phi(\theta, \eta_2) = \theta q_u \exp\{-\Lambda_u(T_c^* + \eta_2 + T_u + \delta)\},$$

$$B(\theta, \eta_2) = \theta - A(\theta, \eta_2) - \Phi(\theta, \eta_2).$$

Конкретные значения θ и η , а также обозначения вероятностей переходов для каждого возможного типа состояний даны в табл. 7.5 (здесь $\beta_0 = 1 - \lambda - \mu$). Заметим, что для любого из состояний $(i, k > 1, 1)$ (яркие эллипсы на рис. 7.22) вероятность опроса θ должна определяться в зависимости от способа, которым было достигнуто это состояние. Если ОТ перешла в такое состояние из $(i, k - 1, 0)$, тогда $\theta = \rho_d$, так как ОТД может быть в любом состоянии перед этим переходом; в противном случае $\theta = \omega = 1 - \exp\{-\Lambda_d T_c^*\}$, так как ОТД была пуста за цикл до перехода. Чтобы учесть эту особенность и сохранить признаки модели Маркова, каждое из этих состояний расщепляется на два подсостояния $(i, k, 1_0)$ и $(i, k, 1_1)$, достигаемые из $(i, k - 1, 0)$ и $(i, k - 1, 1)$ соответственно. Очевидно, $(i, 1, 1)$ состоит только из $(i, 1, 1_0)$.

Таблица 7.5. Значения θ и η и обозначения вероятностей переходов

i, k, l	Значения θ и η				Обозначения				
	θ	η_0	η_1	η_2	α	β	φ	ψ	γ
$0, 1, l > 1$	1	—	—	$t^P(\rho_d)$	λ	β_0	μ	—	—
$i, k < W_i - 1, 0$	ρ_d	$t_0^P(\xi_1)$	—	—	—	—	—	ψ_0	γ_0
$i, W_i - 1, 0$	ρ_d	$t^P(\xi_1)$	—	—	—	—	—	ψ_0^0	γ_0^0
$i, W_i, 0$	1	$t^P(\rho_d)$	—	—	—	—	—	λ^0	γ^*
$i, k < W_i - 1, l > 1$	ω	—	$t_0^P(\omega)$	$t^P(\xi_2)$	α_2	β_2	φ_2	ψ_2	γ_2
$i, W_i - 1, l > 1$	ω	—	$t^P(\omega)$	$t^P(\xi_2)$	α_2	β_2	φ_2	ψ_2^0	γ_2^0
$i, W_i, l > 1$	1	—	—	$t^P(\xi_2^*)$	α_2^*	β_2^*	φ_2^*	—	—
$i, k < W_i - 1, 1_0$	ρ_d	—	$t_0^P(\omega)$	$t^P(\xi_1^*)$	α_1	β_1	φ_1	ψ_1	γ_1
$i, k < W_i - 1, 1_1$	ω	—	$t_0^P(\omega)$	$t^P(\xi_2)$	α_2	β_2	φ_2	ψ_2	γ_2
$i, W_i - 1, 1_0$	ρ_d	—	$t^P(\omega)$	$t^P(\xi_1^*)$	α_1	β_1	φ_1	ψ_1^0	γ_1^0
$i, W_i - 1, 1_1$	ω	—	$t^P(\omega)$	$t^P(\xi_2)$	α_2	β_2	φ_2	ψ_2^0	γ_2^0
$i, W_i, 1_0$	1	—	—	$t^P(\rho_d)$	λ	β_0	μ	—	—
$i, W_i, 1_1$	1	—	—	$t^P(\xi_2^*)$	α_2^*	β_2^*	φ_2^*	—	—

Теперь можно определить вероятность ρ_u^1 непустой ОТ j при условии $v_j \notin X$:

$$\rho_u^1 = 1 - \sum_{i=0}^I \sum_{k=1}^{W_i} \pi_u(i, k, 0) / \left\{ 1 - \sum_{i=1}^I \sum_{k=2}^{W_i} \left[\pi_u(i, k, 1_1) + \sum_{l=2}^k \pi_u(i, k, l) \right] \right\}.$$

Кроме того, в табл. 7.5 $t_0^p(\omega) = \omega(T_d + \delta)$, а вероятности ξ (с различными индексами) того, что противоположная ОТД не будет пустой перед следующим опросом, равны:

$$\xi_1 = \rho_d \left[1 - \frac{q_d \pi_d(1)}{\rho_d} \exp\{-\Lambda_d(T_c^* + T_d + 2\delta + t_0)\} \right] + (1 - \rho_d)\omega,$$

$$\xi_1^* = 1 - \frac{q_d \pi_d(1)}{\rho_d} \exp\{-\Lambda_d(T_c^* + T_d + 2\delta + T_u)\},$$

$$\xi_2 = 1 - q_d \exp\{-\Lambda_d(T_c^* + T_d + T_u + 2\delta)\},$$

$$\xi_2^* = 1 - \exp\{-\Lambda_d(2T_c^* + t_0 + T_u + 2\delta)\} - \omega(1 - \xi_2).$$

Теперь можно найти стационарные вероятности $\pi_u(i, k, l)$. Чтобы упростить вычисления, будем использовать $\pi_u^*(i, k, l) = \pi_u(i, k, l) / \pi_u(0, 1, 0)$ вместо $\pi_u(i, k, l)$. Очевидно, $\pi_u^*(0, 1, 0) = H_0 = 1$, тогда как для $i > 0$

$$\pi_u^*(i, k < W_i, 0) = H_{i-1} \gamma^*(\gamma_0)^{k-1}, \quad (7.28)$$

$$\pi_u^*(i, W_i, 0) = H_i = H_{i-1} \gamma^* \gamma_0^0 (\gamma_0)^{W_i-2}, \quad (7.29)$$

$$\pi_u^*(i, 1, 1_0) = H_{i-1} \lambda^0, \quad (7.30)$$

$$\pi_u^*(i, 1 < k < W_i, 1_0) = H_{i-1} \psi_0 \gamma^*(\gamma_0)^{k-2}, \quad (7.31)$$

$$\pi_u^*(i, W_i, 1_0) = H_{i-1} \psi_0^0 \gamma^*(\gamma_0)^{W_i-2}, \quad (7.32)$$

$$\pi_u^*(i, 1 < k < W_i, 1_1) = \gamma_1 \pi_u^*(i, k-1, 1_0) + 1(k > 2) \gamma_2 \pi_u^*(i, k-1, 1_1),$$

$$\pi_u^*(i, W_i, 1_1) = \gamma_1^0 \pi_u^*(i, W_i-1, 1_0) + 1(W_i > 2) \gamma_2^0 \pi_u^*(i, W_i-1, 1_1),$$

$$\pi_u^*(i, k < W_i, 2) = \psi_1 \pi_u^*(i, k-1, 1_0) + 1(k > 2) [\psi_2 \pi_u^*(i, k-1, 1_1) + \gamma_2 \pi_u^*(i, k-1, 2)],$$

$$\pi_u^*(i, W_i, 2) = \psi_1^0 \pi_u^*(i, W_i-1, 1_0) + 1(W_i > 2) [\psi_2^0 \pi_u^*(i, W_i-1, 1_1) + \gamma_2^0 \pi_u^*(i, W_i-1, 2)],$$

$$\pi_u^*(i, 2 < k < W_i, 2 < l \leq k) = 1(k > l) \gamma_2 \pi_u^*(i, k-1, l) + \psi_2 \pi_u^*(i, k-1, l-1),$$

$$\pi_u^*(i, W_i, 2 < l \leq W_i) = 1(W_i > l) \gamma_2^0 \pi_u^*(i, W_i-1, l) + \psi_2^0 \pi_u^*(i, W_i-1, l-1).$$

[Здесь 1 (условие) — булев оператор, равный единице, если условие выполнено.] Когда $i = I$, правые части соотношений (7.28)–(7.32) должны быть поделены на $1 - \gamma^* \gamma_0^0 (\gamma_0)^{W_I-2}$.

Для состояний $(0, 1, l > W_I + 1)$ легко показать, что

$$\pi_u^*(0, 1, l) = \pi_u^*(0, 1, W_I + 1) (\lambda/\mu)^{l-W_I-1}$$

и их сумма равна

$$S_\infty = \sum_{l=W_I+1}^{\infty} \pi_u^*(0, 1, l) = \pi_u^*(0, 1, W_I + 1) / \left(1 - \frac{\lambda}{\mu}\right).$$

Наконец, $\pi_u^*(0, 1, l)$ для $0 < l \leq W_I + 1$ находятся с помощью уравнений глобального баланса, выписанных для $l = 0, \dots, W_I$:

$$\begin{aligned} \mu\pi_u^*(0, 1, 1) &= 1 - \sum_{i=1}^I [\mu\pi_u^*(i, W_i, 1_0) + \varphi_2^*\pi_u^*(i, W_i, 1_1)] - \\ &\quad - \sum_{i=1}^I \sum_{k=1}^{W_i-1} [\varphi_1\pi_u^*(i, k, 1_0) + 1(k > 1)\varphi_2\pi_u^*(i, k, 1_1)], \\ \mu\pi_u^*(0, 1, l) &= (1 - \beta_0)\pi_u^*(0, 1, l-1) - 1(l > 2)\lambda\pi_u^*(0, 1, l-2) - \\ &\quad - \sum_{i=1}^I (A_{il}^* + B_{il}^* + \Phi_{il}^*), \quad l > 1, \end{aligned}$$

где $A_{i2}^* = 0$,

$$A_{i3}^* = \lambda\pi_u^*(i, W_i, 1_0) + \alpha_2^*\pi_u^*(i, W_i, 1_1) + \sum_{k=1}^{W_i-1} \alpha_1\pi_u^*(i, k, 1_0) + \sum_{k=2}^{W_i-1} \alpha_2\pi_u^*(i, k, 1_1),$$

$$A_{il}^* = 1(l \leq W_i + 1) \sum_{k=l-2}^{W_i-1} \alpha_2\pi_u^*(i, k, l-2) + 1(l \leq W_i + 2)\alpha_2^*\pi_u^*(i, W_i, l-2)$$

для $l > 3$,

$$B_{il}^* = 1(l \leq W_i) \sum_{k=l-1}^{W_i-1} \beta_2\pi_u^*(i, k, l-1) + 1(l \leq W_i + 1)\beta_2^*\pi_u^*(i, W_i, l-1)$$

для $l > 2$; B_{i2}^* определяется подобно A_{i3}^* с заменой λ , α_2^* , α_1 и α_2 на β_0 , β_2^* , β_1 и β_2 ,

$$\Phi_{il}^* = 1(l < W_i) \sum_{k=l}^{W_i-1} \varphi_2\pi_u^*(i, k, l) + 1(l \leq W_i)\varphi_2^*\pi_u^*(i, W_i, l).$$

Суммируя все $\pi_u^*(i, k, l)$, находим стационарные вероятности:

$$\pi_u(i, k, l) = \pi_u^*(i, k, l) / \sum_{(j, h, m)} \pi_u^*(j, h, m).$$

Фактически, стационарные вероятности вычисляются итеративно: используя начальные значения T_{sl}^{d0} , T_{sl}^{d*} и T_{sl}^{d1} , вычисляем вероятности переходов и стационарные вероятности сначала для модели ОТД, а затем для модели ОТ. Наконец, находим модифицированные значения T_{sl}^{d0} , T_{sl}^{d*} и T_{sl}^{d1} и используем полусуммы модифицированных и начальных значений в качестве новых начальных значений. Вычисления прекращаются, когда абсолютные разности между двумя последовательными значениями T_{sl}^{d0} , T_{sl}^{d*} и T_{sl}^{d1} становятся меньше некоторого малого порога.

7.6.3 Оценка показателей производительности

Оценим сначала среднее время обслуживания пакета на МАС-уровне. Начнем с пакетов, переданных после ожидания в очереди (категория 1), т. е. со случая,

когда либо пакет не является первым в группе, либо группа, к которой принадлежит пакет, приходит в непустую очередь. Вторая категория состоит из пакетов, переданных без ожидания в очереди. Очевидно, среднее время обслуживания пакета первой категории равно $D_1^d = T_c^* + T_{sl}^{d1}$ для ОТД и $D_1^u = T_c^* + T_{sl}^{u1}$ для ОТ.

Теперь рассмотрим пакеты второй категории. Для ОТД, используя основное допущение, находим, что среднее время обслуживания такого пакета равно $D_2^d = (T_c^* + T_{sl}^{d0})/2 + T_{sl}^{d1}$. Средние количества всех пакетов и пакетов второй категории, прибывающих в данную ОТД за цикл, равны соответственно $n^d = \Lambda_d T_c$, где

$$T_c = T_c^* + (1 - \rho_d) \left[T_{sl}^{d0} + \frac{\lambda_d^0}{\mu_d^*} T_{sl}^{d*} \right] + \left[\rho_d - \frac{(1 - \rho_d)\lambda_d^0}{\mu_d^*} \right] T_{sl}^{d1}$$

— средняя длительность цикла, и $n_2^d = \Lambda_d q_d (1 - \rho_d) (T_c^* + T_{sl}^{d0})$. Поэтому пакет принадлежит второй категории с вероятностью

$$\kappa_2^d = n_2^d / (\Lambda_d T_c^d) = q_d (1 - \rho_d) (T_c^* + T_{sl}^{d0}) / T_c, \quad (7.33)$$

а искомое среднее время обслуживания для ОТД равно

$$M_d = (1 - \kappa_2^d) D_1^d + \kappa_2^d D_2^d. \quad (7.34)$$

Для ОТ можно использовать похожую формулу: $M_u = (1 - \kappa_2^u) D_1^u + \kappa_2^u D_2^u$, где требуется найти вероятность пакетов второй категории κ_2^u и их среднее время обслуживания D_2^u . Запишем их в следующем виде:

$$\kappa_2^u = q_u \kappa_0 / T_c, \quad D_2^u = \frac{1}{\kappa_0} \sum_{i=0}^I \sum_{k=1}^{W_i} s_{ik0} D_{ik} \pi_u(i, k, 0),$$

где $\kappa_0 = \sum_{i=0}^I \sum_{k=1}^{W_i} s_{ik0} \pi_u(i, k, 0)$. Здесь s_{ikl} и D_{ik} — средняя продолжительность цикла (i, k, l) , в начале которого ОТ находится в состоянии (i, k, l) , и среднее время обслуживания пакета, приходящего в цикле $(i, k, 0)$.

Найдем s_{ikl} и D_{ik} :

$$s_{ikl} = T_c^* + 1(l > 0)[\theta_{ikl}(T_u + \delta + \eta_2^{ikl}) + (1 - \theta_{ikl})\eta_1^{ikl}] + 1(l = 0)[\theta_{ikl}(t_0 + \delta) + \eta_0^{ikl}],$$

где θ_{ikl} , η_0^{ikl} , η_1^{ikl} и η_2^{ikl} определяются в соответствии с табл. 7.5, и

$$D_{ik} = \frac{s_{ik0}}{2} + T_u + \delta + t_{ACK} + (1 - \rho_d)\Delta_{ik},$$

где $\Delta_{ik} = F_D(W_i - k)$ при $k < W_i$, $\Delta_{i, W_i} = F_D(W_{i+1})$ при $i < I$ и $\Delta_{I, W_I} = F_D(W_I)$;

$$F_D(W) = 1(W > 2)\omega \sum_{j=0}^{W-3} [(j+1)T_c^* + T_d + \delta](1 - \omega)^j + \\ + 1(W > 1)(1 - \omega)^{W-2}[(W-1)T_c^* + t^p(\xi_1)],$$

t_{ACK} — среднее время до получения подтверждения, что происходит (в большинстве случаев) в течение опроса следующего терминала, чей слот ненулевой. Таким образом,

$$t_{ACK} = [\rho_d T_d + (1 - \rho_d)\nu_p t_0] / [\rho_d + (1 - \rho_d)\nu_p].$$

Теперь оценим средние времена (T_d^{MAC} и T_u^{MAC}) пребывания пакетов в ОТД и ОТ. Эти показатели находятся из формулы Литтла: $T_d^{MAC} = q_d L_d / \Lambda_d$ и $T_u^{MAC} = q_u L_u / \Lambda_u$, и основная задача состоит в оценке средних размеров ОТД (L_d) и ОТ (L_u), измеренных в пакетах. Начнем с оценки L_d . Очевидно, $L_d = S_L^d / T_c$, где

$$S_L^d = \sum_l \bar{l}_d(l) t_c^d(l) \pi_d(l),$$

$$\bar{l}_d(l) = \frac{l}{q_d} + \frac{\Lambda_d}{2q_d} [T_c^* + T_{sl}^d(l)] - 1(l > 0) \left[1 - \frac{T_{sl}^d(l) - \delta}{T_c^* + T_{sl}^d(l)} \right],$$

а $t_c^d(l) = T_c^* + T_{sl}^d(l)$. Окончательно имеем:

$$\begin{aligned} S_L^d = & \pi_d(0) \frac{\Lambda_d}{2q_d} (T_c^* + T_{sl}^{d0})^2 + \\ & + \pi_d(1) \left\{ \frac{T_c^* + T_{sl}^{d*}}{q_d} \left[1 - q_d + \frac{\Lambda_d}{2} (T_c^* + T_{sl}^{d*}) \right] + T_{sl}^{d*} - \delta \right\} + \\ & + \frac{\pi_d(2) D_1^d}{q_d(1 - \lambda_d/\mu_d)} \left(l_1^d + 1 + \frac{1}{1 - \lambda_d/\mu_d} \right), \end{aligned}$$

где $l_1^d = \frac{\Lambda_u}{2} D_1^d - q_d \left[1 - \frac{T_{sl}^{d1} - \delta}{D_1^d} \right]$. Для ОТ используем аналогичную формулу: $L_u = S_L^u / T_c$, где

$$S_L^u = \sum_{(i,k,l)} \bar{l}_u(i,k,l) s_{ikl} \pi_u(i,k,l),$$

$$\bar{l}_u(i,k,l) = \frac{l}{q_u} + \frac{\Lambda_u}{2q_u} s_{ikl} - 1(l > 0) \theta_{ikl} \left[1 - \frac{T_u + \delta + t_{ACK}}{T_c^* + T_u + \delta + \eta_2^{ikl}} \right].$$

После преобразований получаем:

$$L_u = (q_u^{-1} D_1^u \kappa_1^* + \kappa_0^*) / T_c,$$

где

$$\kappa_1^* = \sum_{l=1}^{W_I} (l + l_u^1) \pi_u(0, 1, l) + (l_u^1 + W_I) S_\infty + \pi_u(0, 1, W_I + 1) / \left(1 - \frac{\lambda}{\mu} \right)^2,$$

$$l_u^1 = \frac{\Lambda_u}{2} D_1^u - q_u \left(1 - \frac{T_u + \delta + t_{ACK}}{D_1^u} \right),$$

$$\kappa_0^* = s_{010} \frac{\Lambda_u}{2q_u} \pi_u(0, 1, 0) + \sum_{(i,k,l): i>0} \bar{l}_u(i,k,l) s_{ikl} \pi_u(i,k,l).$$

7.6.4 Численные результаты

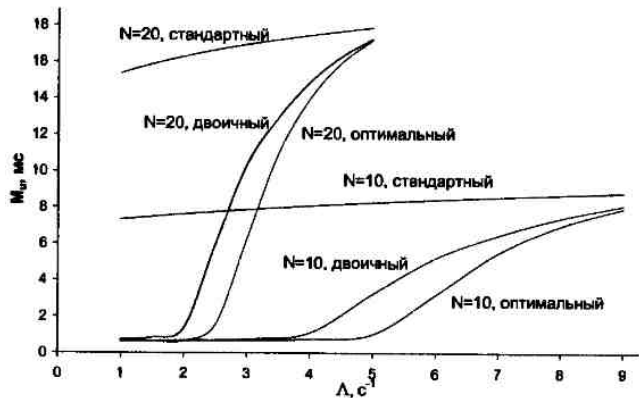
Применим разработанный аналитический метод для оценки производительности протокола PCF в зависимости от параметров трафика (нагрузки) и конфигурации сети (т.е. от числа терминалов N), а также для сравнения алгоритмов стандартного опроса (СО), двоичной схемы (ДС) с $I = 8$ и $W_i = 2^i$, предложенной в [135], и оптимального динамического опроса (ОДО). Вид последней схемы определяется путем применения аналитического метода для нахождения

оптимальных параметров $(I, \{W_i\})$, обеспечивающих минимальное значение показателя M_u или T_u^{MAC} для каждой точки пространства $\Omega = (\Lambda_d, \Lambda_u, q_d, q_u, N)$. Таким образом, схема ОДО требует слежения за изменениями параметров восходящего и нисходящего трафика и оперативной коррекции параметров $(I, \{W_i\})$.

Подавляющая часть трафика, проходящего через беспроводную сеть, связана с работой стека протоколов ТСП/ІР, для которого характерно равенство интенсивностей поступления пакетов восходящего и нисходящего трафика, так как каждому ТСП-пакету (средняя длина которого принята равной 576 байтам, что характерно для многошаговых ТСП-соединений [80]) соответствует ТСП-подтверждение (его длину считаем равной 80 байтам). Поэтому в данном численном исследовании предположим, что $\Lambda_d = \Lambda_u = \Lambda$ и $q_d = q_u = q$. Кроме того, применим следующие распределения длины пакетов L : $L = 576$ и $L = 80$ с вероятностями 0,7 и 0,3 для ОТД и с вероятностями 0,3 и 0,7 для ОТ, что примерно соответствует случаю, когда треть соединений реализует восходящий ТСП-трафик.

Начнем с анализа зависимости среднего времени обслуживания M_u и M_d пакетов восходящего и нисходящего трафика от нагрузки на сеть соответственно. На рис. 7.23 и 7.24 показаны зависимости M_u и M_d от Λ при $q = 0,1$, $N = 10$ и $N = 20$ для трех исследуемых схем опроса, причем оптимальный режим будем искать, минимизируя M_u для $I = 1$ при варьировании W_1 от 2 до $W_{\max} = 256$. Графики зависимости оптимального значения $W_1 = W_1^{opt}$ от Λ приведены на рис. 7.25 (сплошные линии). Сразу заметим, что увеличение числа стадий I в схеме ОДО не позволяет существенно улучшить производительность: например, для случая $N = 20$, $q = 0,1$ и $\Lambda = 3,5 \text{ с}^{-1}$ при $I = 1$ оптимальным окном является $W_1^{opt} = 35$, обеспечивающее $M_u^{\min} = 6,3 \text{ мс}$, а при $I = 3$ имеем те же самые $W_1^{opt} = W_2^{opt} = W_3^{opt} = 35$. Поэтому далее исследуется только ОДО с единственной стадией опроса ($I = 1$).

Рис. 7.23. Зависимость среднего времени обслуживания M_u пакетов в ОТ от нагрузки Λ на сеть при $q = 0,1$ для различного числа терминалов N и различных схем опроса



Вернемся к графикам на рис. 7.23 и 7.24. Видно, что при ненасыщенных очередях сети обе схемы динамического опроса значительно лучше стандартного опроса. В частности, при низкой нагрузке и большом числе терминалов динамический опрос позволяет на порядок снизить среднее время обслуживания по

сравнению со стандартным опросом. Сравним теперь ДС и ОДО. При средних нагрузках ОДО существенно превосходит ДС, т. е. среднее время обслуживания при ОДО в два-три раза меньше. Однако при низкой нагрузке не так уж важно, какой именно способ динамического опроса использовать (при одинаковых максимально возможных окнах), так как любой терминал почти все время находится на стадии I с $W_i = W_I$. При больших нагрузках обе динамические схемы приближаются к СО, так как при наличии непустых ОТД и ОТ терминалы опрашиваются в каждом цикле, как при стандартном опросе. Заметим также, что оптимальное окно имеет тенденцию к уменьшению при увеличении нагрузки (см. рис. 7.25), а при низких нагрузках принимает максимально возможное значение.

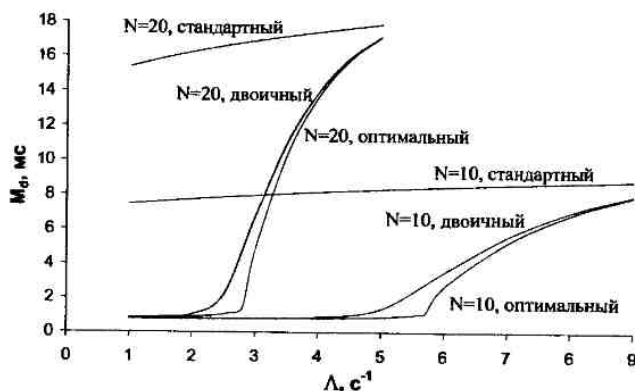


Рис. 7.24. Зависимость среднего времени обслуживания M_d пакетов в ОТД базовой станции от нагрузки Λ на сеть при $q = 0,1$ для различного числа терминалов N и различных схем опроса

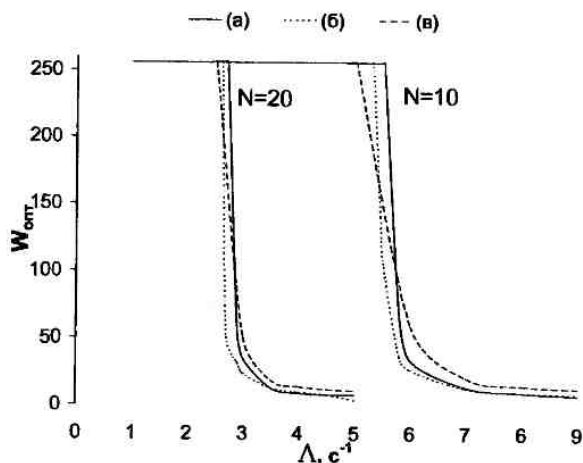


Рис. 7.25. Зависимость оптимального окна опроса W_{opt} от нагрузки Λ при оптимизации а — M_u , б — T_u^{MAC} и в — $T_u^{MAC} + T_u^{MAC}$

Рассмотрим теперь, какое влияние оказывает групповой характер трафика на среднее время обслуживания пакетов восходящего трафика. На рис. 7.26 представлен случай $N = 10$. Кривые на рис. 7.26 разделены на две группы в зависимости от параметра $\Lambda/q = \text{const}$, характеризующего интенсивность прибытия

пакетов. Эти графики также показывают, что динамический опрос всегда лучше стандартного, а ОДС обеспечивает существенно меньшее значение $M_u^{\min}$ при средних значениях q и Λ/q . При больших Λ/q разница в производительности анализируемых схем значительно меньше (20–30%). Кроме того, оказалось (см. также рис. 7.23), что в отличие от динамического опроса среднее время обслуживания при СО весьма слабо зависит как от Λ и q , так и от их отношения.

Рис. 7.26. Среднее время обслуживания в зависимости от группового характера трафика: а — $\Lambda/q = 50$, б — $\Lambda/q = 80$ и в — $\Lambda/q = 30$ при различных схемах опроса: СО — стандартный опрос, ДС — двоичная схема, ОДО — оптимальный динамический опрос

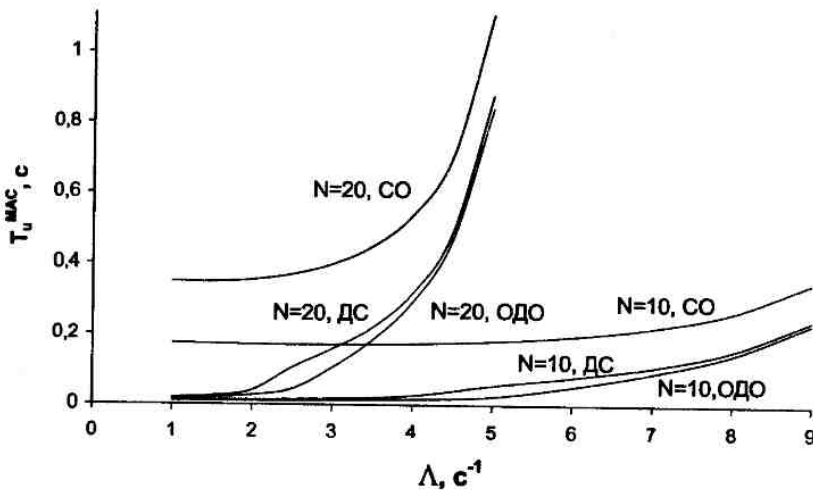
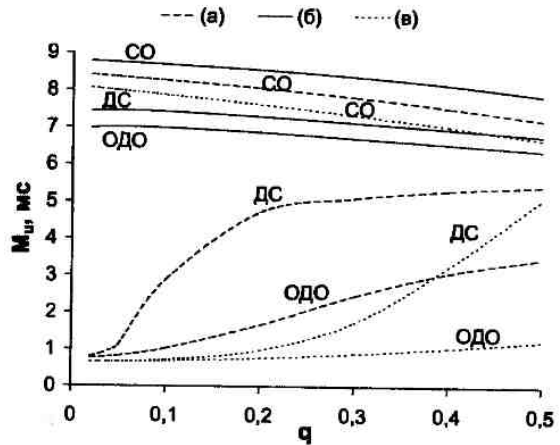


Рис. 7.27. Зависимость среднего времени T_u^{MAC} пребывания пакетов в ОТ от нагрузки при различных схемах опроса

Рассмотрим теперь другие показатели производительности — средние времена T_u^{MAC} и T_d^{MAC} пребывания пакетов в ОТ и ОТД соответственно. Снова будем сравнивать три типа опроса: стандартный, универсальный двоичный с $I = 8$, $W[1] = 2^i$ и оптимальный, но оптимизировать теперь будем не M_u , а T_u^{MAC} . На рис. 7.27–7.29 показаны графики зависимости T_u^{MAC} и T_d^{MAC} от параметра Λ

и T_u^{MAC} от параметра q , аналогичные рис. 7.25, 7.24 и 7.26 (соответствующие зависимости $W_1^{opt}(\Lambda)$ даны на рис. 7.25 пунктиром). Как и следовало ожидать, значения времен пребывания пакетов в очередях при трех анализируемых схемах соотносятся примерно так же, как и значения времен обслуживания.

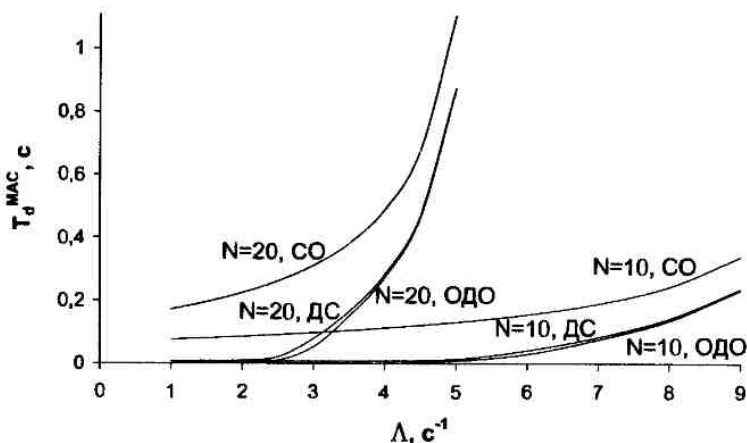


Рис. 7.28. Зависимость среднего времени T_d^{MAC} пребывания пакетов в ОД от нагрузки при различных схемах опроса

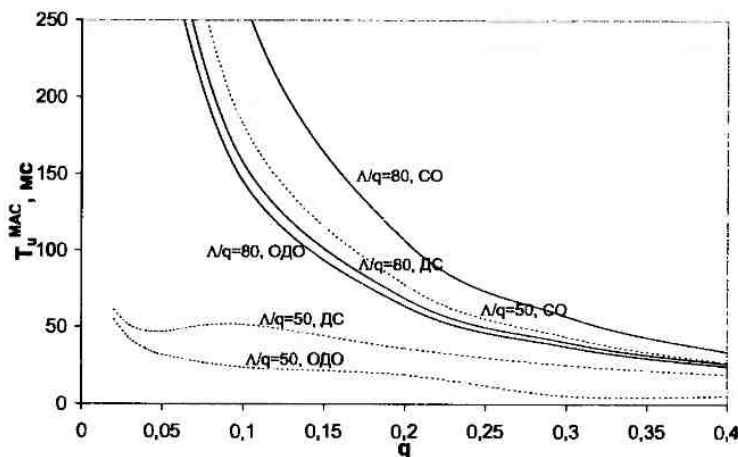


Рис. 7.29. Средние времена пребывания пакетов в ОД в зависимости от группового характера трафика

Также интересным показателем производительности сетей с ТСП-трафиком является сумма $T_u^{MAC} + T_d^{MAC}$, так как именно эта сумма равна времени нахождения ТСП-сегмента (представляющего собой сначала ТСП-пакет, а затем — его подтверждение) в МАС-очередях беспроводной сети. Это время может являться определяющей компонентой важнейшего параметра протокола ТСП, так называемого Round Trip Time — среднего времени от отправления пакета источником до получения им подтверждения. Учитывая похожий вид зависимостей

$T_u^{MAC}(\Lambda)$ и $T_d^{MAC}(\Lambda)$ (см. рис. 7.27 и 7.28), нетрудно предугадать форму кривых $T_u^{MAC} + T_d^{MAC}$ в зависимости от Λ , данных на рис. 7.30 для $N = 10$ и $N = 20$.

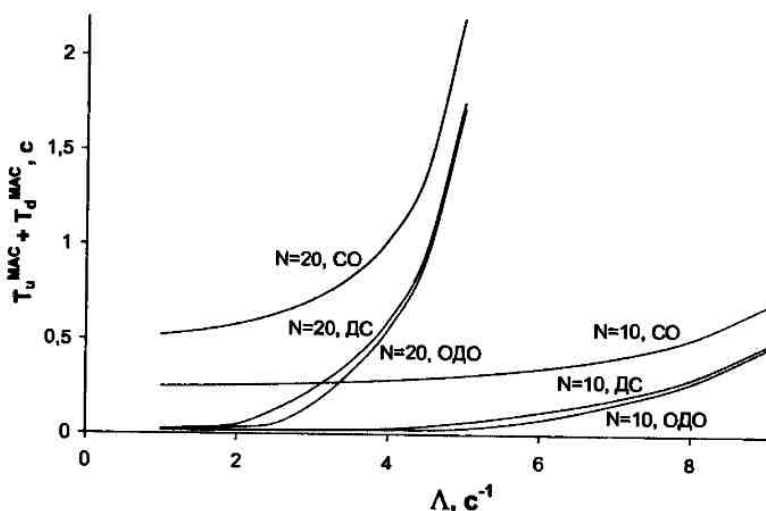


Рис. 7.30. Зависимость среднего времени нахождения TCP-сегмента в MAC-очередях беспроводной сети от нагрузки

Исследуем теперь, как отличаются оптимальные режимы, найденные при минимизации M_u , T_u^{MAC} и $T_u^{MAC} + T_d^{MAC}$ (сплошные, пунктирные и штриховые линии на рис. 7.25). На первый взгляд, кажется, что разница довольно велика, но на самом деле выбор критерия оптимизации оказывается не так существенен. Рисунок 7.31 демонстрирует, что разница значений $M_u^{\min}$ и M_u^* при окне W_1^{opt} , полученном путем минимизации соответственно M_u и T_u^{MAC} , не превышает 5%, а в большинстве случаев $M_u^{\min}$ и M_u^* практически совпадают.

В заключение данного раздела следует отметить, что предложенная стратегия динамического опроса и методы его моделирования обещают быть полезными и для других беспроводных протоколов, таких как IEEE 802.15 (Bluetooth) [135] и IEEE 802.16 (см. главу 8).

7.7 Экспериментальные исследования и оптимизация городской беспроводной сети

В данном разделе исследуется беспроводная сеть передачи данных Radionet [14, 25], разработанная и реализованная Институтом проблем передачи информации РАН (ИППИ РАН) для подключения к сети Интернет организаций науки и образования г. Москвы. Указанная сеть надежно функционирует в г. Москве в течение шести лет, обеспечивая доступ в Интернет порядка 60 организаций науки и образования. Основным техническим средством, на котором реализована данная радиосеть, является беспроводной мост Cisco Aironet 340 Series Wireless Bridge, работающий по протоколу IEEE 802.11b [161]. Сеть Radionet имеет сетовую струк-

туру (рис. 7.2), причем ее базовые станции размещены на высотных зданиях и соединены между собой, а также с ИППИ РАН оптоволоконными линиями связи.

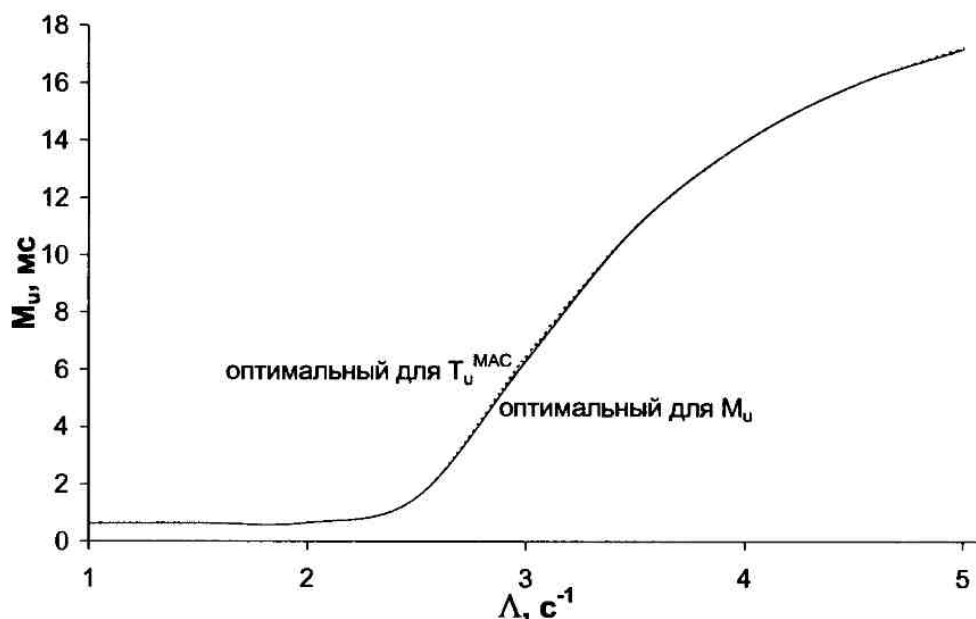


Рис. 7.31. Сравнение оптимальных режимов в зависимости от способа их нахождения для $N = 20$, $q = 0,1$.

В целом как в России, так и за рубежом ведутся интенсивные научные исследования беспроводных радиосетей на основе протокола IEEE 802.11, направленные на повышение их эффективности и выбор оптимальных параметров протокола. В имеющихся работах оценка производительности таких сетей проводится либо путем имитационного моделирования, либо с помощью приближенных аналитических моделей, однако вопрос о получении исходных данных для этих моделей остается открытым. Для решения данной проблемы в ИППИ РАН в настоящее время разрабатывается система сбора и анализа реальных данных о работе сети Radionet, о динамических характеристиках трафика, проходящего через эту сеть, и частотах искажения пакетов из-за радиопомех.

Для исследования была выбрана одна из сот, базовая станция которой расположена на высотном здании на Кудринской площади. Эта базовая станция обслуживает пять станций, каждая из которых является порталом для своей локальной сети. Расстояния от базовой станции до конечных станций велики (несколько километров) и различны. Антенны конечных станций сфокусированы на базовую и не принимают сигналы друг от друга.

Для решения задачи выбора оптимальных параметров работы сети, а также получения исходных данных для построения аналитических моделей сетей проведено исследование, направленное на сбор статистики работы реальной беспроводной сети и анализ полученных данных в целях выявления особенностей ее функционирования.

7.7.1 Способы получения и усреднения статистических данных

Способ получения исходных данных. В данном исследовании изучаются следующие показатели работы сети: интенсивность передаваемого и принимаемого базовой станцией трафика, количество принимаемых и передаваемых пакетов и количество повторных передач пакетов. Эти показатели выбраны из всей совокупности предоставляемых беспроводным мостом данных как наиболее полно отражающие характер функционирования беспроводной сети. Указанные данные периодически снимаются в течение долгого времени с беспроводного моста Cisco Aironet 340 Series Wireless Bridge, являющегося базовой станцией. Аналогичная информация не собирается с оконечных станций, так как иначе собранные данные передавались бы через радиоканалы, искажая статистику.

В качестве основного средства получения данных использован протокол SNMP. Протокол SNMP относится к семейству систем управления и мониторинга. Назначение таких систем — организация взаимодействия между менеджером и агентом, где в качестве менеджера обычно выступает некоторая прикладная программа, управляющая некоторым подключенным устройством посредством агента, располагающегося на этом устройстве.

Функцию сбора исходных данных выполняет специально разработанная программа-менеджер. Программа написана на языке Perl и исполняется на персональном компьютере, работающем под управлением операционной системы Linux и подключенном к беспроводному мосту по проводным линиям связи. Основной функцией программы является периодический запрос и занесение в базу данных следующих SNMP-переменных:

- `statThruRfRxBytesTotal` и `statThruRfTxBytesTotal` — полное число байт, успешно принятых и переданных беспроводным мостом;
- `statThruRfRxPktTotal` и `statThruRfTxPktTotal` — количество пакетов, успешно принятых и переданных беспроводным мостом;
- `statRadRetries` — количество раз, когда пакет пришлось посылать заново из-за потери либо самого пакета, либо подтверждения о его приеме.

* Программа работает в течение долгого времени и использует для работы общие каналы. Поэтому для уменьшения загрузки этих каналов период считывания выбран достаточно большим — 10 секунд. Этот период чересчур велик, чтобы проследить за отдельными событиями, такими как прохождение пакета или появление ошибки, однако вполне достаточен для исследования интегральных характеристик работы сети. Считанные данные заносятся в базу данных, которая реализована на СУБД MySQL и представляет собой таблицу, содержащую времена снятия параметров и их значения.

Для получения интегральных статистических данных была написана программа-обработчик. В ее функции входит чтение записей базы данных, относящихся к определенным периодам времени, вычисление реальных значений параметров работы, исходя из MIB-переменных, усреднение параметров по времени и вычисление оцениваемых характеристик трафика и сети.

Способы усреднения статистических данных. Как известно, процессы в телекоммуникационной сети носят стохастический характер. Как утверждается в работе [169], эти процессы относятся к классу self-similar, или самоподобных процессов. Особенностью этих процессов является медленное, полиномиальное убывание функции автокорреляции, что выражается в сильных биениях графика, изображающего изменения параметра процесса, причем биения не пропадают при усреднениях по большому промежутку времени. В дальнейшем планируется проверить гипотезы о характере трафика и ошибок. Цель же данного раздела несколько иная: исследовать зависимость характеристик работы сети от различных факторов, таких как время суток, рабочий или выходной день. Основной интересующей нас характеристикой является среднее значение того или иного параметра, например интенсивность трафика. Очевидно, что это значение будет существенным образом меняться в зависимости от времени суток, например ночью оно будет существенно меньше, чем днем. С другой стороны, обнаруживаются большие случайные колебания этой интенсивности, если интервал усреднения выбран слишком коротким. Поэтому, чтобы свести случайные колебания к минимуму, но не упустить из виду особенности, связанные с суточным изменением характеристик, выбран интервал усреднения, равный одному часу.

Вычисление интенсивности помех. Искажения передаваемых кадров в общем случае происходят в результате воздействия как радиопомех, так и коллизий, причем принимающая станция не может определить конкретную причину искажений, что является особенностью протокола IEEE 802.11. В рассматриваемой сети оконечные станции скрыты друг от друга, что приводит к большой вероятности коллизий передаваемых ими кадров. Кадры, передаваемые базовой станцией, искажаются в результате коллизии только в случае одновременного начала передачи с базовой станцией и с оконечной станцией — станцией назначения, но этот случай можно считать очень маловероятным. Таким образом, почти все искажения пакетов, передаваемых базовой станцией, происходят только из-за городских радиопомех. Этот факт используется для оценки интенсивности радиопомех — Bit Error Rate, BER.

Оценка BER производится в предположении, что биты искажаются независимо друг от друга с равной вероятностью, которая медленно меняется со временем. По определению BER, кадр длиной l бит искажается с вероятностью $p_l^* = 1 - \exp(-l * \text{BER})$. Очевидно, значение BER мало и поэтому приблизительно равно вероятности искажения одного бита. Способ снятия статистики не позволяет точно определить длину конкретного пакета, поэтому будем считать, что все кадры, содержащие пакеты, имеют одинаковую длину l , равную средней длине за исследуемый интервал времени. Тогда p_l^* можно оценить как $p_l^* = \frac{R}{R+N}$, где R — число пакетов, на которые не было получено подтверждение, т. е., пренебрегая вероятностью искажения кадра подтверждения приема на MAC уровне (ACK-кадра), число повторов передачи. Значение R соответствует значению переменной `statRadRetries`, снимаемой с беспроводного моста, а N — число успешно дошедших пакетов, находимое из переменной `statThruRfTxPktTotal`. Искомое BER выражается из предыдущих формул и равно:

$$\text{BER} = -\ln\left(\frac{N}{R+N}\right)/l,$$

при этом средняя длина l выражается следующим образом:

$$l = L_H + 8 * \frac{N_b}{N},$$

где N_b — число удачно переданных байт, которое соответствует значению переменной `statThruRfTxBytesTotal`, L_H — число бит заголовка MAC- и PHY-уровней, определенное протоколом [222].

Видно, что BER вычисляется при множестве предположений, поэтому результат следует воспринимать как оценочный. Основную погрешность вносит предположение о том, что все пакеты имеют одинаковую длину. На самом деле, конечно, длины пакетов различны, причем более длинные пакеты имеют большую вероятность исказиться и поэтому будут чаще встречаться среди повторяемых. К сожалению, распределение длин пакетов невозможно описать каким-либо простым законом, так как оно зависит от используемых сетевых протоколов. Например, при передаче файла по протоколу TCP он будет разбит на множество пакетов, имеющих одинаковую длину, и один — последний пакет — будет иметь меньшую длину. Таким образом, чтобы оценить распределение хотя бы в этом случае, придется сделать предположение о распределении длин файлов. В реальной сети данные передаются многими способами, поэтому итоговое распределение получает причудливый вид, меняющийся от часа к часу.

Еще одним фактором, искажающим результат, является сомнительность допущения о независимом характере искажения бит. Как утверждается в работе [236], радиопомехи носят характер длительных всплесков, затрагивающих множество бит, а возможно, и множество пакетов. На данном этапе не представляется возможным подтвердить или опровергнуть это предположение, так как за 10-секундный период снятия статистики через эфир успевают пройти сотни пакетов. Искажение нескольких пакетов подряд из-за всплеска помех и искажение такого же количества пакетов из-за случайного изменения отдельных бит для нас неотличимы. В дальнейшем будут рассмотрены возможности проверки данной гипотезы, а пока решено остановиться на предположении о независимых искажениях отдельных бит данных.

7.7.2 Анализ полученных данных

Основная характеристика функционирования сети — это количество переданной в секунду информации, называемое пропускной способностью или трафиком. Для выяснения зависимости интенсивности трафика от времени суток значения были усреднены по всем дням, для которых имелась статистика, отдельно для рабочих дней и отдельно для выходных, так как очевидно, что характер работы в рабочие и в выходные дни существенно различается. Ниже приведены графики зависимости трафика от времени суток (рис. 7.32).

Из графиков видно, что передаваемый трафик в среднем существенно больше принимаемого — это значит, что сеть используется по большей части для получения данных из сети Интернет. Ясно видно повышение нагрузки в рабочие дни с 8 до 21 часа.

Провал в районе 12 часов, вероятно, вызван перерывом на обед с 12 до 13 часов. В выходные, как и ожидалось, не наблюдается существенного повышения

нагрузки в дневные часы. Кроме того, трафик на прием и трафик на передачу практически одинаковы, и не обнаруживают значительных отличий от ночных значений, что позволяет сделать предположение о том, что этот трафик вызван работой сетевых роботов, например устройств, реализующих протоколы маршрутизации.

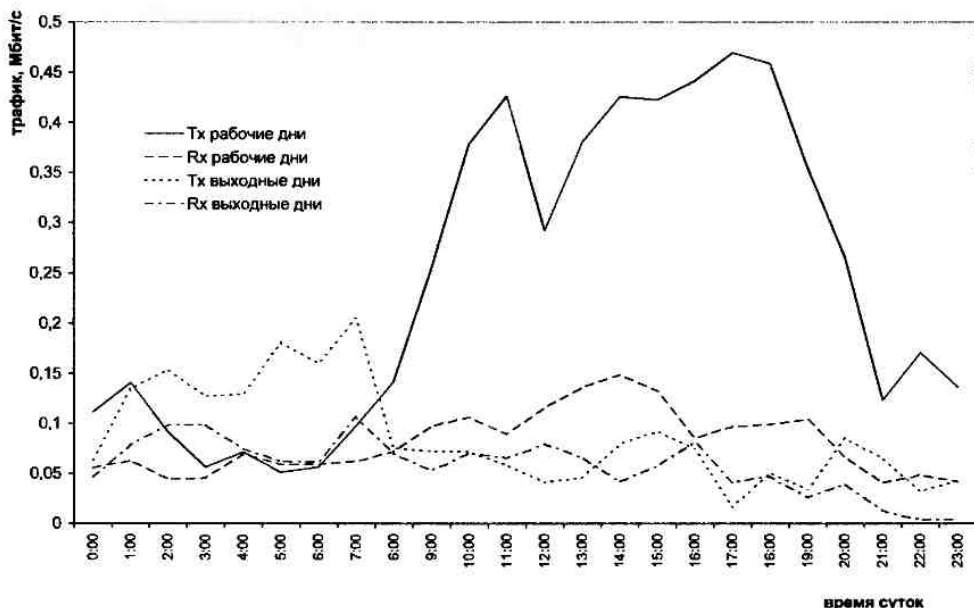


Рис. 7.32. Интенсивность трафика (Тх-передача, Rx-прием)

Характер работы сетевых протоколов становится более ясен при рассмотрении зависимости числа пакетов от времени (рис. 7.33).

Отчетливо видно, что количество переданных и принятых пакетов практически полностью совпадает, причем в случае передачи от базовой станции к оконечной — это длинные пакеты с данными, а в случае передачи от оконечной станции к базовой — короткие пакеты-подтверждения. Это позволяет сделать вывод, что в нашей сети в основном используется протокол TCP/IP, в котором на каждый переданный пакет должно поступить подтверждение от станции, которая этот пакет приняла.

Еще одна характеристика функционирования сети — распределение длин пакетов. На данном этапе оказалось возможно получить лишь оценочные результаты. Дело в том, что исходными данными для обработки служили количества байт и пакетов, прошедших через беспроводной мост за 10-секундный интервал времени. Следовательно, разделив одно на другое, можно узнать лишь среднюю величину интересующего нас показателя за 10 секунд. Ниже (рис. 7.34) приведены графики распределений длин пакетов, принимаемых и передаваемых базовой станцией. Видно, что базовая станция передает в основном длинные пакеты: пакеты с данными, получаемыми из внешней сети, а принимает короткие: пакеты с TCP-подтверждениями.

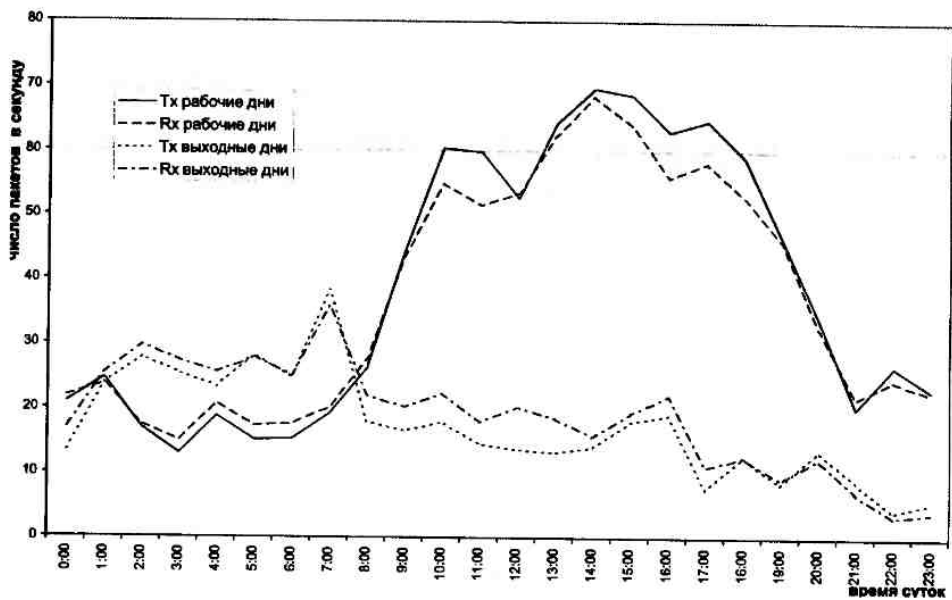


Рис. 7.33. Прохождение пакетов (Tx-передача, Rx-прием)

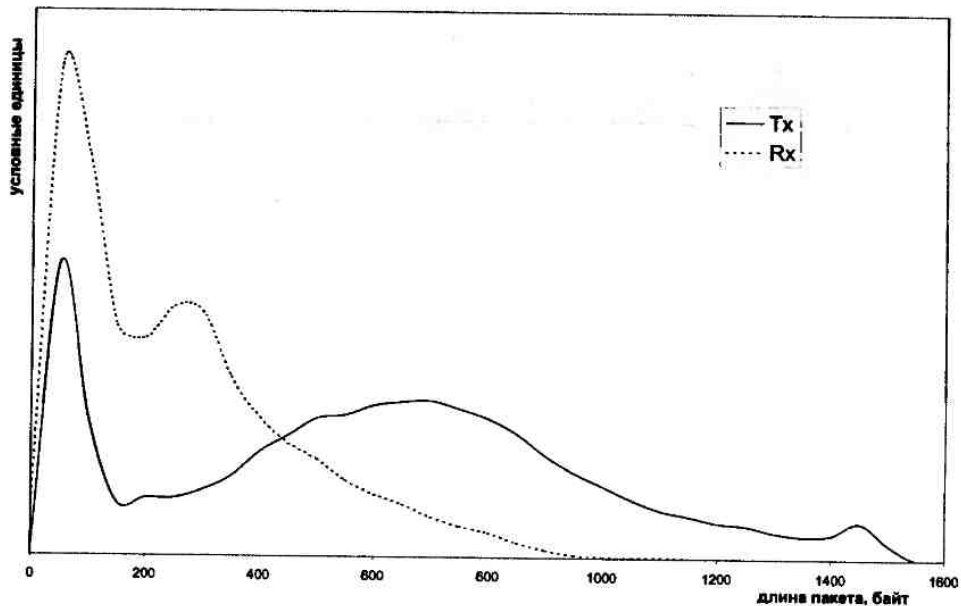


Рис. 7.34. Распределения длин пакетов

Приведенные выше зависимости отражают нагрузку на сеть, которая зависит по большому счету от активности пользователей и подвержена сильным недетерминированным флуктуациям. Напротив, такой параметр, как интенсивность радиопомех (BER), меньше зависит от человеческого фактора (см. рис. 7.35).

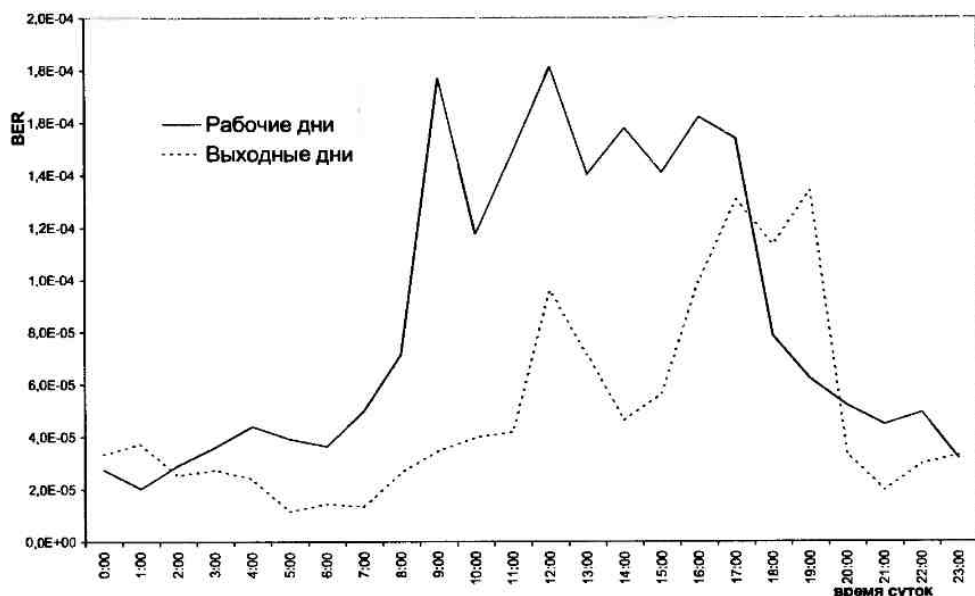


Рис. 7.35. Интенсивность радиопомех (BER)

Как и следовало ожидать, в рабочее время уровень помех существенно повышается. Тот факт, что пик спадает к 18 часам, а также то, что рост помех в выходные менее выражен, позволяет сделать вывод, что помехи в основном вызваны как взаимным влиянием других базовых станций, так и работой «частотных пиратов», т. е. организаций, предоставляющих беспроводной доступ в сеть Интернет, не имея на это соответствующих лицензий и работающих в том же частотном диапазоне, что и ИППИ РАН.

7.7.3 Постановка задачи оптимизации

Эффективность работы беспроводной сети во многом зависит от того, как много теряется (а значит, и посылается повторно) пакетов. Как уже было отмечено, вероятность искажения пакета, посылаемого базовой станцией, в основном зависит только от уровня радиопомех в эфире. Этот уровень описывается величиной SNR (Signal-to-Noise Ratio, отношение сигнал/шум), которая задается известной формулой

$$\text{SNR} = 20 \log_{10} \left(\frac{V_s}{V_n} \right),$$

где V_s и V_n — уровни сигнала и шума соответственно, или величиной BER (Bit Error Rate) — вероятностью того, что один бит информации исказится при передаче. В случае некодированной передачи зависимость между BER и SNR определяется типом модуляции сигнала

$$\text{BER} = Q[\text{SNR}],$$

где $Q[\cdot]$ — некоторый нелинейный оператор. В случае кодированной передачи реальный уровень $\text{BER}(V_i, \text{SNR})$ получается намного ниже по причине того, что передаваемые символы, искаженные помехами, на приемной стороне могут быть восстановлены декодером приемника. Поэтому BER проще и нагляднее представлять с помощью опытных данных и построенных по ним графиков зависимостей $\text{BER}(V_i, \text{SNR})$. Далее для сокращения обозначений вместо $\text{BER}(V_i, \text{SNR})$ будем писать просто BER, подразумевая его зависимость от V_i и от SNR.

Вероятность искажения пакета PER_i (Packet Error Rate) для какой либо из скоростей передачи V_i равна

$$\text{PER}_i(l) = 1 - \prod_{k=1}^l [1 - \text{BER}_k],$$

а в предположении независимости искажения отдельных бит вычисляется по формуле

$$\text{PER}_i(l, \text{BER}) = 1 - (1 - \text{BER})^l,$$

где l — длина пакета в битах. Это предположение выполняется лишь приближенно. В различных модуляционных схемах кодируются сразу несколько бит, поэтому от искажений тех или иных символов или комбинаций символов зависит искажение тех или иных бит. Влияние этих искажений друг на друга можно учесть, вводя для каждой схемы кодирования поправку α_i к BER, которая выражает усредненную по всевозможным кодовым векторам степень влияния искаженных символов на закодированные биты. В этом случае предыдущая формула примет вид

$$\text{PER}_i(l, \text{BER}) = 1 - (1 - \alpha_i \cdot \text{BER}(V_i, \text{SNR}))^l, \quad (7.35)$$

где индекс i обозначает номер скорости V_i , для которой производятся вычисления. Приближенное значение поправочного коэффициента α_i может быть вычислено из зависимостей $\text{PER}_i(l, \text{BER})$, снятых для длинных пакетов ($l \approx 1000$ байт). В таких пакетах встречаются всевозможные сочетания бит, а искажения символов происходят достаточно часто и случайным образом. Поэтому оценка α_i для длинных пакетов близка к среднему значению α_i , взятому по всем возможным кодовым векторам и всем возможным их искажениям. То есть данная оценка α_i близка к математическому ожиданию $M[\alpha_i]$.

Наконец, учитывая тот факт, что в сетях IEEE 802.11b заголовок физического уровня длиной h_{phy} бит передается в основном на минимальной скорости V_1 , эта формула приобретает следующий вид:

$$\text{PER}_i(l, \text{BER}) = 1 - (1 - \alpha_1 \cdot \text{BER}(V_1, \text{SNR}))^{h_{phy}} (1 - \alpha_i \cdot \text{BER}(V_i, \text{SNR}))^l,$$

где l — длина MAC-части кадра.

При известной интенсивности потерь пакетов $\text{PER}_i(l, \text{BER})$ среднее количество повторных передач $N_i(l, \text{BER})$ (предполагая, что их предел неограничен) нефраgmentированного пакета длины l на скорости V_i находится по формуле

$$N_i(l, \text{BER}) = \frac{\text{PER}_i(l, \text{BER})}{1 - \text{PER}_i(l, \text{BER})},$$

где $P_i(l, \text{BER}) = 1 - (1 - \text{PER}_i(l, \text{BER}))(1 - \text{PER}_i(l_{\text{ACK}}, \text{BER}))$ — вероятность неудачной передачи пакета длиной l из-за искажения либо кадра DATA, либо от-

ветного АСК-кадра длиной $l_{\text{АСК}}$. Отсюда можно найти среднее время $T_i(l, \text{BER})$ передачи пакета длины l .

Так как вероятность искажения пакета при передаче увеличивается при увеличении его длины, мы должны знать, каковы длины передаваемых в нашей сети пакетов. Это нужно для вычисления среднего времени передачи пакета $T_i^{\text{CP}}(\text{BER})$, которое находится по формуле

$$T_i^{\text{CP}}(\text{BER}) = \sum_l \Lambda(l) \cdot T_i(l, \text{BER}), \quad (7.36)$$

где $\Lambda(l)$ — вероятность появления пакета длины l , а суммирование производится по всем возможным значениям длин, для которых $\Lambda(l) \neq 0$. На рис. 7.34 представлена зависимость вероятности появления в трафике нашей сети пакетов от их длины.

Задача данного исследования состоит в улучшении эффективности схемы DCF в части пропускных способностей, обеспечиваемых для клиентов, в условиях помех, характерных для города и искажающих передаваемые пакеты. Задача может быть решена посредством двух механизмов.

Первым механизмом является фрагментация, т.е. разбиение передаваемых пакетов на фрагменты максимальной длины L_f , с тем чтобы в случае искажения помехами повторно передавать не весь пакет целиком, а только его фрагмент. Чем меньше длина фрагмента, тем меньше вероятность его искажения при передаче, но в то же время, с каждым фрагментом посылается его заголовок длиной L_h и принимается подтверждение его успешного приема (АСК-кадр), повышается время бесполезного занятия среды.

Вторым механизмом является выбор скорости передачи путем применения различных типов модуляции сигнала и схем кодирования. Проведенные ранее исследования [133] показали, что интенсивность радиопомех (BER) зависит не только от отношения сигнал/шум (SNR), но и от способа модуляции сигнала, а значит, и скорости передачи. Вид данной зависимости, полученной по экспериментальным данным [133], показан на рис. 7.36.

Итак, задача состоит в разработке метода выбора оптимальных порога фрагментации и скорости передачи для конкретной сети на основе полученной ранее статистики. Порог фрагментации и скорость будут оптимальными, если при передаче пакета станция будет как можно меньше занимать среду доступа, т.е. необходимо минимизировать среднее время передачи пакета T_{CP} без учета пренебрежимо малого времени распространения сигнала.

Здесь сразу возникают некоторые трудности, а именно станция «не знает» текущих величин BER и SNR, поэтому требуется ввести некий параметр, по которому станция будет судить о необходимости переключения скорости и/или применения того или иного порога фрагментации. В качестве такого параметра можно взять, например, оценки тех же BER и SNR. Однако эти оценки не доступны для прямого измерения и требуется найти такой параметр, который был бы, во-первых, «известен» базовой станции и в то же время его можно было однозначно связать с вероятностью искажения кадра и со средним временем передачи кадра при оптимальном пороге фрагментации. Во-вторых, текущее значение данного параметра не должно зависеть от распределения длин пакетов и от интегральных параметров, таких как средняя длина кадра, среднее время передачи кадра и т. д.

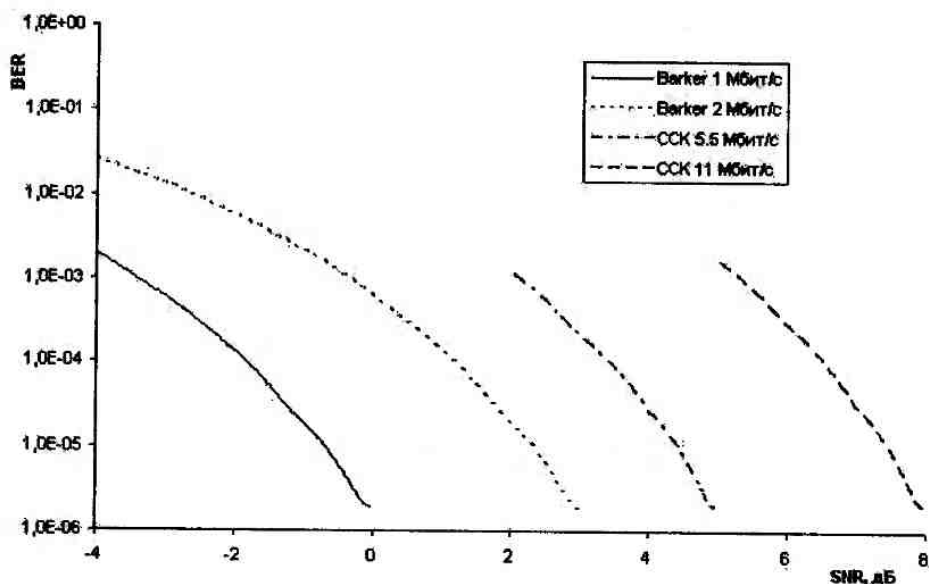


Рис. 7.36. Зависимость интенсивности радиопомех от отношения сигнал/шум для разных типов модуляций, применяемых в IEEE 802.11b

Было принято решение в качестве такого параметра взять количество повторных передач пакетов в расчете на один удачно дошедший пакет $N^{ret} = R/N$. При таком выборе параметра выполняются все указанные выше требования. Единственным требованием, которое накладывается на величины R и N , является $\min\{R, N\} \gg 1$. Оно нужно для обеспечения необходимой точности вычислений.

7.7.4 Метод оптимизации

Пусть L — длина пакета, а L_f — некоторый выбранный порог фрагментации, тогда количество фрагментов $n(L) + 1$ и длина последнего фрагмента L_0 находятся из формул

$$n(L) = \lfloor (L - 1)/L_f \rfloor, \quad L_0 = L - L_f \cdot n(L).$$

Отсюда время передачи всего пакета находится как сумма времен передачи отдельных фрагментов

$$T_i(L, \text{BER}) = n(L) \cdot T_i(L_f, \text{BER}) + T_i(L_0, \text{BER}).$$

Если при отсутствии искажений пакет длины L передается за время

$$T_i(L, 0) = n(L) \cdot [t_h + t_{L_f} + t_{ACK} + 2\text{SIFS}] + t_h + t_{L_0} + t_{ACK} + \text{SIFS} + \text{DIFS},$$

то при наличии искажений тот же пакет передается уже за время

$$T_i(L, \text{BER}) = n(L) \left[t_h + t_{L_f} + t_{ACK} + 2\text{SIFS} + \frac{P'_i(L_f)}{1 - P'_i(L_f)} t_{bad}(L_f) \right] + \\ + t_h + t_{L_0} + t_{ACK} + \text{SIFS} + \text{DIFS} + \frac{P'_i(L_0)}{1 - P'_i(L_0)} t_{bad}(L_0) + \sum_{j=1}^{n(L)+1} \langle \tau_B \rangle_j.$$

Здесь:

$t_h, t_{L_f}, t_{L_0}, t_{ACK}$ — времена передачи соответственно заголовка физического уровня, фрагментов и кадра ACK;

$t_{bad}(L_f) = t_h + t_{L_f} + \text{EIFS}$ — время неудачной попытки передачи фрагмента длиной L_f ;

$P'_i(L_m) = P_i(L_h + L_m, \text{BER})$ — вероятность неудачной передачи фрагмента размером L_m на скорости V_i и уровне радиопомех BER;

$\langle \tau_B \rangle_j$ — усредненное время отсрочки при передаче j -го фрагмента.

Найдем $\langle \tau_B \rangle_j$. Так как $\tau_B = (w - 1) \cdot R(0, 1) \cdot \sigma$, где $R(0, 1)$ — случайное число, выбранное из интервала $(0, 1)$, а σ — слот отсрочки, то среднее значение τ_B равно $\tau_B' = \frac{1}{2} \cdot \sigma \cdot (w - 1)$. Учитывая все возможные повторные передачи одного фрагмента, получаем $\langle \tau_B \rangle_j$:

$$\langle \tau_B \rangle_j = \frac{1}{2} \cdot \sigma \cdot \left[W_0 \cdot \sum_{k=k_0}^m [P_j^*]^k \cdot 2^k + W_m \cdot \sum_{k=m+1}^{\infty} [P_j^*]^k - \sum_{k=k_0}^{\infty} [P_j^*]^k \right],$$

где $k + 1$ — номер попытки передачи (предполагается неограниченное число попыток), а также:

при $n(L) = 0$: $k_0 = 0$ и $P_1^* = P'_i(L)$;

при $n(L) > 0$: $k_0 = 0$ при $j = 1$ и $k_0 = 1$ при $j > 1$, $P_j^* = P'_i(L_f)$ при $j \leq n(L)$ и $P_{n(L)+1}^* = P'_i(L_0)$.

Для первого передаваемого фрагмента среднее время отсрочки рассчитано из тех соображений, что пакет, поступающий на передачу, практически всегда попадает в очередь вследствие занятости среды и/или наличия в очереди на передачу других пакетов. Это происходит по причине высокой загрузки исследуемой базовой станции (особенно в будни в рабочие часы), что подтверждается опытными данными. Поэтому передача первого фрагмента начинается с отсчета времени отсрочки исходя из конкурентного окна $w = W_0$. Вычисляя последовательно все суммы, входящие в выражение для $\langle \tau_B \rangle_j$ по формуле геометрической прогрессии, окончательно получаем:

$$\langle \tau_B \rangle_j = \frac{1}{2} \cdot \sigma \cdot \left[\frac{W_0 \cdot 2^{k_0} \cdot [P_j^*]^{k_0} - 2 \cdot W_m \cdot [P_j^*]^{m+1}}{1 - 2P_j^*} + \frac{W_m \cdot [P_j^*]^{m+1} - [P_j^*]^{k_0}}{1 - P_j^*} \right].$$

Так как пакеты имеют различные длины L , то и передаваться они будут за различные времена $T_i(L, \text{BER})$. Нас интересует среднее время передачи пакета, которое находится по той же формуле (7.36). Находя для каждого значения скорости V_i и отношения сигнал/шум SNR минимум $T_i^{\text{cp}}(\text{BER})$ по L_f , получаем зависимости оптимального порога фрагментации и времени передачи от SNR для конкретного распределения длин пакетов $\Lambda(L)$.

Количество повторных передач $N_i^{\text{ret}}(\text{BER})$ в расчете на один пакет можно найти по следующей формуле:

$$N_i^{\text{ret}}(L, \text{BER}) = n(L) \cdot \frac{P_i(L_f, \text{BER})}{1 - P_i(L_f, \text{BER})} + \frac{P_i(L - n(L) \cdot L_f, \text{BER})}{1 - P_i(L - n(L) \cdot L_f, \text{BER})}.$$

Далее, для оптимальных порогов фрагментации вычисляем среднее количество повторных передач

$$N_i^{ret}(\text{BER}) = \sum_L \Lambda(L) \cdot N_i^{ret}(L, \text{BER}).$$

Вычисление этого показателя необходимо, если требуется создать режим работы станции, при котором она автоматически переключает скорость и/или подстраивает порог фрагментации под уровень радиопомех. Естественно, используя предложенный выше метод, станция сможет отследить только медленные изменения среднего уровня радиопомех, так как ей необходимо в течение некоторого времени (порядка нескольких секунд) накапливать статистику, оценивающую количество повторных передач в расчете на один пакет $N_i^{ret}(\text{SNR})$.

7.7.5 Результаты расчетов

Анализируя зависимости $\text{PER}_i(l, \text{BER})$, снятые для длинных пакетов ($l \approx 1000$ байт) и приведенные в [133], и $\text{BER}(\text{SNR})$ (см. рис. 7.36), были вычислены величины поправочных коэффициентов α_i для формулы (7.35). Оказалось, что

$$\alpha_1 = \alpha_2 = 8/11, \quad \alpha_3 = 7/16, \quad \alpha_4 \approx 0,35,$$

где индекс i при α_i , как и ранее, обозначает скорость передачи: 1 Мбит/с для $i = 1$; 2 Мбит/с для $i = 2$; 5,5 Мбит/с для $i = 3$ и 11 Мбит/с для $i = 4$. Благодаря этому стал возможным расчет вероятностей искажения пакетов $\text{PER}_i(l, \text{BER})$ для любых скоростей передачи и любых уровней шумов (SNR).

В результате программной реализации методов, изложенных в предыдущем разделе, были получены следующие зависимости: оптимального порога фрагментации от условий среды передачи (см. рис. 7.37) для разных типов модуляций, среднего времени передачи пакета и среднего количества повторных передач, имеющих место при наличии оптимальной фрагментации, от SNR и BER (см. рис. 7.38 и 7.39).

В протоколе IEEE 802.11 схема переключения скорости станции не определена. Один из методов выбора скорости передачи состоит в том, что по достижении определенного уровня повторных передач (N_i^{ret} повторов на пакет) станция переключается на более медленную скорость передачи. Очевидно, что уровень N_i^{ret} должен быть также оптимизирован. По результатам совместной оптимизации порога фрагментации и уровня переключения получена зависимость оптимального порога фрагментации от уровня радиопомех, по которой мы находим оптимальные уровни переключения скоростей (рис. 7.40).

Основная масса беспроводных устройств, работающих на базе протокола IEEE 802.11, не имеет функции определения уровня SNR (например, серия устройств Cisco Aironet 340), поэтому для реализации разработанной схемы переключения скоростей и установки порога фрагментации удобнее использовать среднее значение числа повторов передач на пакет (N_i^{ret}) за некоторый интервал времени (порядка 10 секунд). Это значение определяется исходя из числа пакетов, переданных за этот интервал, и общего числа повторов передач, которые предоставляются интерфейсом этих устройств. На основе этого разработана схема опти-

мального переключения скоростей и установки порога фрагментации в зависимости от уровня повторных передач N_i^{ret} (рис. 7.41).

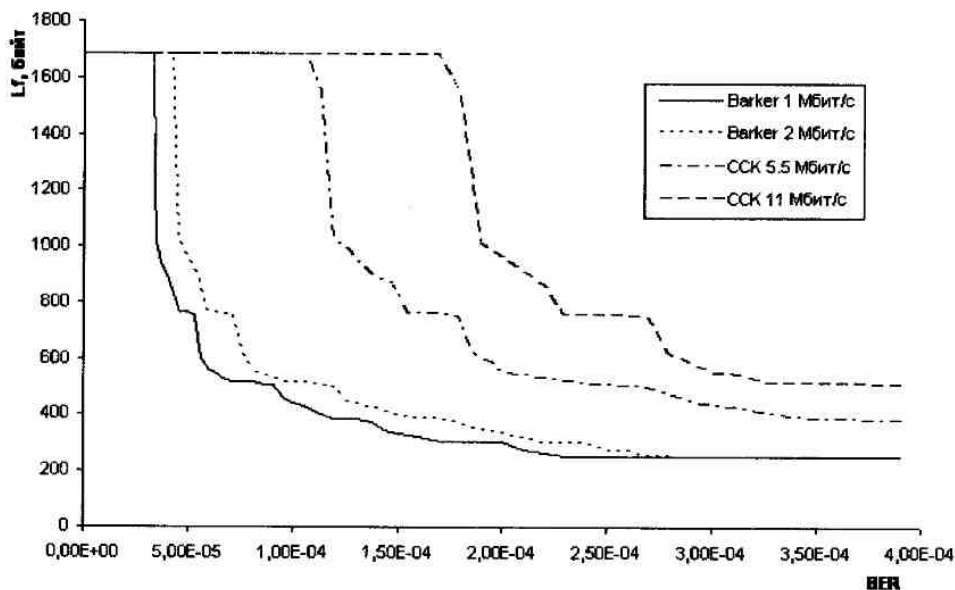


Рис. 7.37. Зависимость оптимального порога фрагментации от уровня радиопомех при различных скоростях передачи

Согласно этой схеме при $0 \leq N^{ret} < 0,3$ станция передает нефрагментированные пакеты на скорости 11 Мбит/с. При достижении $N^{ret} \approx 0,3$ станция понижает скорость до 5,5 Мбит/с и при возрастании N^{ret} от 0,3 до 0,7 включает механизм фрагментации, наконец, при $N^{ret} \approx 0,7$ переключается на скорость 2 Мбит/с с отменой фрагментации. Аналогично, на скорости 2 Мбит/с при $0,2 < N^{ret} < 0,7$ станция включает механизм фрагментации и при $N^{ret} \approx 0,7$ переключается на скорость 1 Мбит/с, затем на скорости 1 Мбит/с при $N^{ret} > 0,15$ устанавливает порог фрагментации в соответствии с зависимостью, показанной на рис. 7.41.

В настоящее время по умолчанию во всех устройствах, работающих по протоколу IEEE 802.11, применяется «однопроцентная» схема, в которой переключения скоростей происходят по 1% переповторов (т. е. $N_i^{ret} = 0,01$). На рис. 7.42 показан результат сравнения эффективности работы полученной оптимальной схемы и применяемой «однопроцентной» схемы. Из этой зависимости ясно видно, что критерий переключения скоростей по 1%-ному уровню повторных передач не является оптимальным — при некоторых значениях отношения сигнал/шум среднее время передачи пакета при оптимальной схеме оказывается в 1,5–2 раза меньше. Заметим также, что в «однопроцентной» схеме вообще не предусмотрена какая-либо фрагментация пакетов.

Из зависимости, представленной на рис. 7.42, также видно, что при малых значениях SNR применение оптимальной фрагментации и переключения в несколько раз снижает среднее время передачи пакета и, следовательно, значительно по-

вышает пропускную способность сети. Все это оправдывает необходимость применения рассматриваемой схемы оптимизации в реальных устройствах.

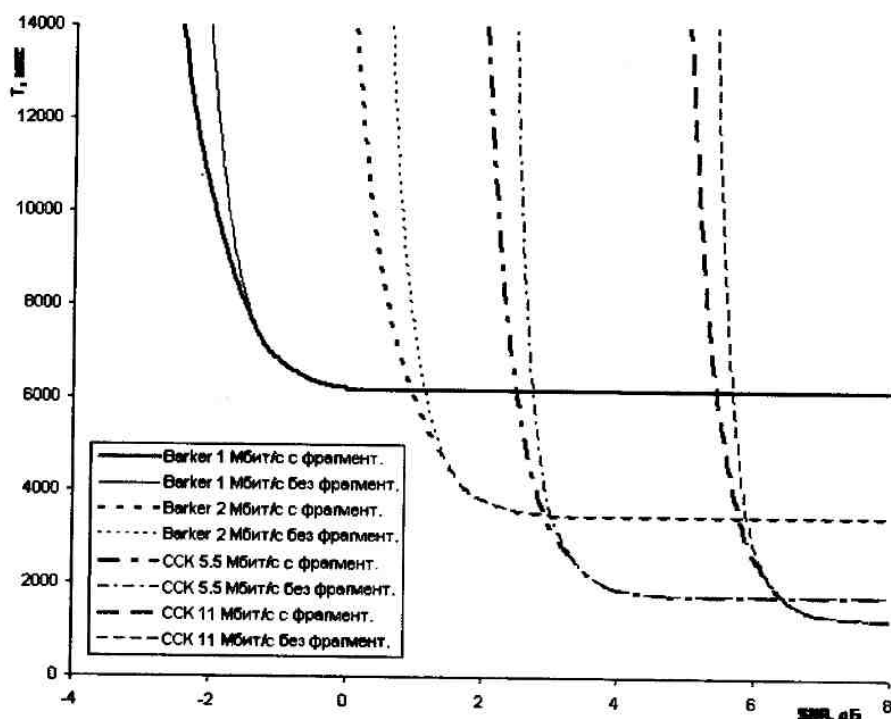


Рис. 7.38. Зависимость среднего времени передачи пакета от уровня шума при различных скоростях передачи (с оптимальной фрагментацией и без фрагментации)

Интересно отметить, что при некоторых значениях SNR , несмотря на большие искажения и большое количество повторно передаваемых пакетов, все же выгодней передавать пакеты, не снижая скорость передачи. Когда высокая скорость передачи перестает компенсировать время, затрачиваемое на повторы передач пакетов, то конечно же станции нужно переключаться на более низкую скорость и заново выбирать порог фрагментации.

Анализ интенсивности радиопомех в сети RadioNet показал, что в рабочее время суток BER достигает значений $1,8 \times 10^{-4}$, а в среднем равен $1,5 \times 10^{-4}$. При этом согласно «однопроцентной» схеме станция передает пакеты на скорости 1 Мбит/с, а отношение сигнал/шум равно $-2,05$ дБ. Тогда среднее время передачи пакета равно 14 500 мкс, в то время как при оптимальной схеме оно равно 10 730 мкс, а порог фрагментации — 335 байтам, при этом процент повторных передач огромен — 80%. Тем не менее среднее время передачи пакета оказывается значительно меньше, чем при использовании «однопроцентной» схемы.

В связи с высокой эффективностью разработанной схемы оптимального переключения скоростей и установки порога фрагментации была создана система мо-

ниторинга и автоматической подстройки параметров беспроводной сети, которая была внедрена в беспроводных региональных сетях ряда частотных операторов России.

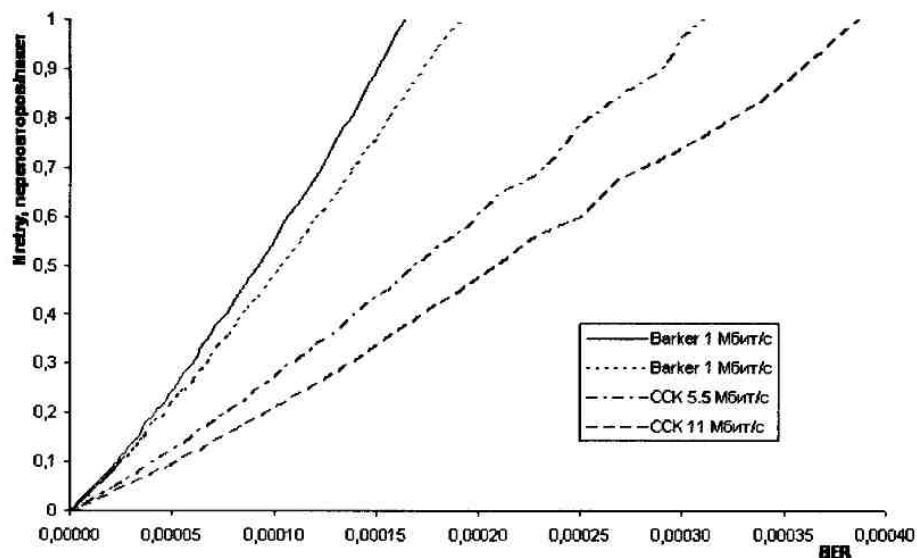


Рис. 7.39. Зависимость среднего количества повторных передач пакета от уровня радиопомех при оптимальной фрагментации для различных скоростей

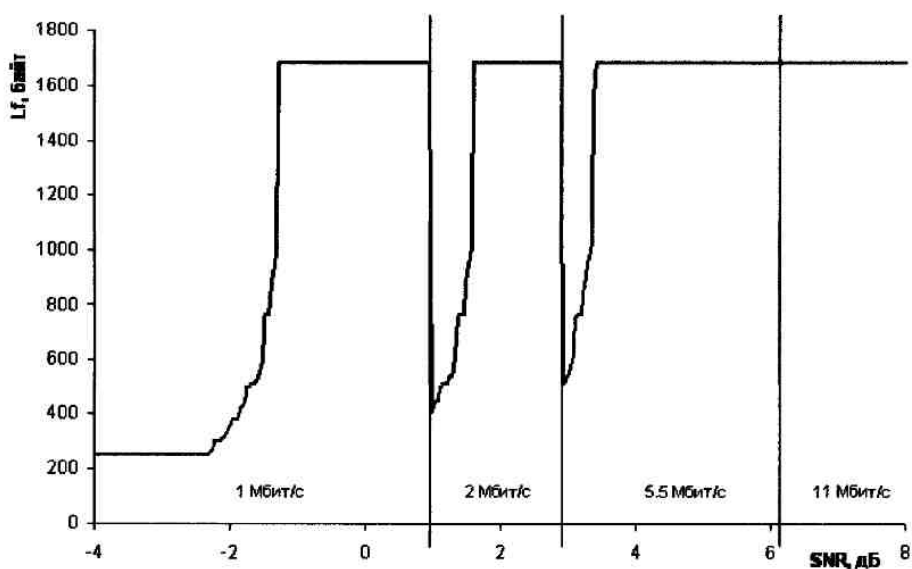


Рис. 7.40. Зависимость оптимального порога фрагментации от уровня радиопомех при оптимальном переключении скоростей

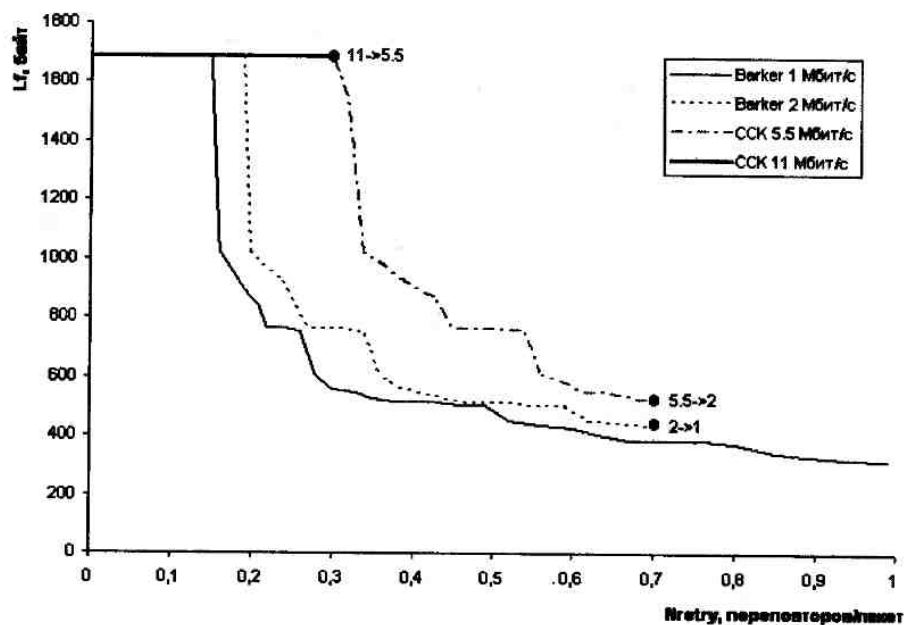


Рис. 7.41. Схема оптимального переключения скоростей и установки порога фрагментации

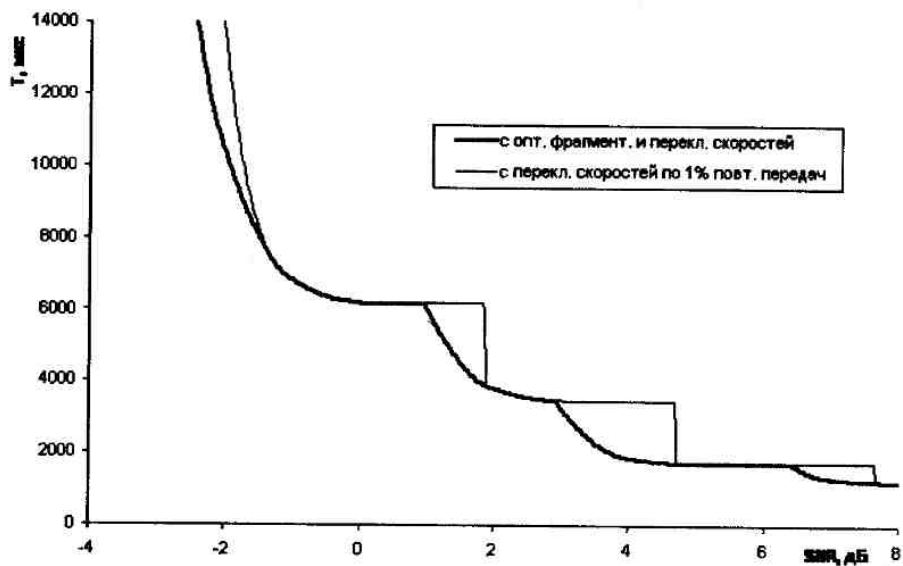


Рис. 7.42. Сравнение среднего времени передачи пакета при оптимальных фрагментации и схеме переключения скоростей и при переключении по 1% повторов передач пакетов

7.8 Математические методы топологического проектирования широкополосных беспроводных сетей

Одной из важнейших задач информатизации Российской Федерации является ликвидация информационного неравенства между регионами. На обширной территории Российской Федерации наряду с информационно развитыми городами и регионами имеются территории со слабой телекоммуникационной инфраструктурой (в первую очередь это касается сельских районов).

Решением проблемы информационного неравенства является либо использование дорогостоящих средств спутниковой связи, либо создание широкополосных беспроводных региональных сетей, проектирование которых является особенно актуальным. Проектирование беспроводных региональных сетей предполагает решение целого ряда сложных взаимосвязанных задач [14]. К ним относятся:

- задача размещения базовых станций,
- оптимальное подключение абонентов региональной сети к базовым станциям (задача оптимального синтеза топологической структуры),
- математические модели и методы оценки производительности и надежности беспроводных сетей,
- выбор оптимальных параметров беспроводных протоколов (типа 802.11x и 802.16x),
- модели взаимодействия беспроводных сетей, работающих по разным протоколам (например, GPRS или 3G с 802.11x и 802.16x) и являющихся частями единой региональной сети.

Вопросам оценки производительности беспроводных сетей были посвящены предыдущие разделы этой главы. В данном разделе описывается и решается задача синтеза топологической структуры региональной сети, включающая оптимальное размещение базовых станций и прикрепление к ним локальных сетей абонентов по критерию минимальной стоимости при наличии ограничений на электромагнитную совместимость.

7.8.1 Постановка задачи оптимального размещения базовых станций и подключения к ним локальных компьютерных сетей абонентов

Формальная постановка задачи имеет следующий вид.

Пусть имеется M мест-кандидатов на размещение станций с известными координатами, N типов станций, K клиентов. Для каждого типа станций известно ограничение на суммарную ширину канала всех подключенных к станции клиентов (производительность): B_n , а также стоимость C_n . Стоимость подключения

клиента к n -й станции:

$$\begin{cases} W_n^1 > 0, & \text{если расстояние между ними } r \leq R_1; \\ W_n^2 > 0, & \text{если расстояние между ними } R_1 < r \leq R_2. \end{cases}$$

При $r > R_2$ подключение невозможно. Для каждого клиента заданы координаты и требуемая ширина канала b_k .

Требуется разместить станции на местах-кандидатах и подключить клиентов к станциям, минимизировав стоимость комплекса.

Введем переменные:

$$X_{mn} = \begin{cases} 1, & \text{если станция } n\text{-й модели размещена} \\ & \text{на } m\text{-м месте;} \\ 0, & \text{в противном случае.} \end{cases}$$

$$Y_{km} = \begin{cases} 1, & \text{если } k\text{-й клиент подключен к станции,} \\ & \text{размещенной на } m\text{-м месте;} \\ 0, & \text{в противном случае.} \end{cases}$$

Пусть также W_{kmn} — стоимость подключения k -го клиента к станции n -го типа при условии, что она размещена на m -ом месте. Если подключение невозможно, считаем $W_{kmn} = -1$.

Потребуем выполнения следующих условий.

1. Каждый клиент должен быть подключен, причем только к одному месту-кандидату:

$$\forall k \quad \sum_{m=1}^M Y_{km} = 1. \quad (7.37)$$

2. На каждом из мест-кандидатов может быть размещено не более одной станции:

$$\forall m \quad \sum_{n=1}^N X_{mn} \leq 1. \quad (7.38)$$

3. На месте-кандидате, к которому подключен клиент, должна быть размещена станция. Согласно (7.37) для заданного $k = \kappa$ из всех элементов $\{Y_{\kappa m}\}_{m=1}^M$ только один отличен от нуля (равняется единице). Пусть этот элемент $Y_{\kappa \mu}$. Из (7.38) следует, что для любого m (в том числе для $m = \mu$) среди элементов $\{X_{mn}\}_{n=1}^N$ не более одного отличны от нуля (равны единице). Для $m = \mu$ нужно, чтобы среди $\{X_{\mu n}\}_{n=1}^N$ обязательно нашелся элемент отличный от нуля. Значит,

$$\forall k \quad \sum_{m=1}^M Y_{km} \left(\sum_{n=1}^N X_{mn} \right) = 1;$$

$$\forall k \quad \sum_{m=1}^M \sum_{n=1}^N Y_{km} X_{mn} = 1. \quad (7.39)$$

4. Для заданного $k = \kappa$ из всех произведений $Y_{km}X_{mn}$ лишь одно отлично от нуля (равняется единице). Сумма $\sum_{m=1}^M \sum_{n=1}^N W_{kmn} Y_{km} X_{mn}$ задает стоимость подключения клиента κ при данной конфигурации сети, определяемой матрицами $\|Y\|_{K \times M}$ и $\|X\|_{M \times N}$ (или показывает, что подключение невозможно, если она отрицательна). Получаем условие:

$$\begin{aligned} \forall k \quad & \sum_{m=1}^M \sum_{n=1}^N W_{kmn} Y_{km} X_{mn} > 0; \\ \forall k \quad & \sum_{m=1}^M \sum_{n=1}^N W_{kmn} Y_{km} X_{mn} \geq 1. \end{aligned} \quad (7.40)$$

Поскольку $W_{kmn} \neq 0$, а при выполнении условий (7.37) и (7.38) среди произведений $Y_{km}X_{mn}$ может быть не более одного отличного от нуля (равного единице), то для любого фиксированного k при выполнении (7.37), (7.38) и (7.40) условие (7.39) является избыточным.

5. а) Если есть возможность назначать трафик (b_k) для каждого клиента, то суммарный трафик всех клиентов, подключенных к станции, размещенной на m -м месте, не должен превышать производительности станции:

$$\forall m \quad \sum_{n=1}^N B_n X_{mn} \geq \sum_{k=1}^K b_k Y_{km}; \quad (7.41)$$

б) Если производительность станции (B_n) делится между подключенными клиентами поровну, то выделенная ширина канала для каждого клиента должна быть не меньше требуемой (b_k):

$$\begin{aligned} \forall m, k \quad & \sum_{n=1}^N B_n X_{mn} \geq b_k Y_{km} \sum_{k'=1}^K Y_{k'm}; \\ \forall m, k \quad & \sum_{n=1}^N B_n X_{mn} \geq \sum_{k'=1}^K b_k Y_{km} Y_{k'm}. \end{aligned} \quad (7.42)$$

Целевая функция (стоимости):

$$\Phi_0 = \sum_{k=1}^K \sum_{m=1}^M \sum_{n=1}^N W_{kmn} Y_{km} X_{mn} + \sum_{m=1}^M \sum_{n=1}^N C_n X_{mn}.$$

Учитывая условие (7.40), из целевой функции можно исключить все слагаемые, в которые в качестве множителя входит $W_{nmk} = -1$:

$$\Phi = \sum_{\substack{k,m,n \\ W_{kmn} > 0}} W_{kmn} Y_{km} X_{mn} + \sum_m \sum_n C_n X_{mn}.$$

Далее решаем задачу $\Phi_0 \rightarrow \min$ с ограничениями (7.37)–(7.39), а также (7.41) или (7.42) при условии, что X_{mn} , Y_{kn} — булевы переменные.

Теорема. Пусть x_j, z — булевы переменные. Тогда справедливо тождество:

$$\begin{cases} \sum_{j=1}^J x_j - z \leq J - 1; \\ -\sum_{j=1}^J x_j + J \cdot z \leq 0 \end{cases} \Leftrightarrow \prod_{j=1}^J x_j = z.$$

Доказательство.

$\Rightarrow$: Пусть

$$\begin{cases} \sum_{j=1}^J x_j - z \leq J - 1; \\ -\sum_{j=1}^J x_j + J \cdot z \leq 0. \end{cases} \quad (7.43)$$

1а) Если $\prod_{j=1}^J x_j = 1$, то $\forall j \quad x_j = 1 \Rightarrow \sum_{j=1}^J x_j = J$, и система (7.43) имеет вид:

$$\begin{cases} J - z \leq J - 1; \\ -J + J \cdot z \leq 0, \end{cases} \Leftrightarrow \begin{cases} z \geq 1; \\ z \leq 0, \end{cases} \Leftrightarrow z = 1.$$

1б) Если $z = 1$, то из системы (7.43) следует:

$$\begin{cases} \sum_{j=1}^J x_j \leq J; \\ \sum_{j=1}^J x_j \geq J; \end{cases} \Leftrightarrow \sum_{j=1}^J x_j = J \Leftrightarrow \forall j \quad x_j = 1 \Leftrightarrow \prod_{j=1}^J x_j = 1.$$

Таким образом, при выполнении (7.43): $\prod_{j=1}^J x_j = z$.

$\Leftarrow$: Пусть $\prod_{j=1}^J x_j = z$.

2а) $\prod_{j=1}^J x_j = z = 0$. Тогда $0 \leq \sum_{j=1}^J x_j \leq J - 1$, и неравенства (7.43) выполняются.

2б) $\prod_{j=1}^J x_j = z = 1$. Тогда $\sum_{j=1}^J x_j = J$, и система (7.43) верна.

Теорема доказана.

Поставим в соответствие каждому произведению $Y_{km}X_{mn}$ переменную Z_{kmn} . Учитывая, что $Y_{km}Y_{k'm} = Y_{k'm}Y_{km}$, в соответствие этим произведениям можно поставить переменную $U_{\max\{k, k'\} \min\{k, k'\} m}$. Таким образом, можно преобразовать задачу к виду:

$$\Phi \equiv \sum_{\substack{k, m, n \\ W_{kmn} > 0}} W_{kmn} Z_{kmn} \sum_m \sum_n C_n X_{mn} \rightarrow \min$$

с ограничениями:

$$\left\{ \begin{array}{l} \forall m, n, k \quad X_{mn} + Y_{km} - Z_{kmn} \leq 1; \\ \forall m, n, k \quad -X_{mn} - Y_{km} + 2Z_{kmn} \leq 0; \\ \forall k \quad \sum_m Y_{km} = 1; \\ \forall m \quad \sum_n X_{mn} \leq 1; \\ \forall k \quad \sum_m \sum_n W_{kmn} Z_{kmn} \geq 1; \\ \forall m \quad \sum_n B_n X_{mn} \geq \sum_k b_k Y_{km} \text{ либо выполняется система ограничений,} \\ \text{приведенная ниже:} \end{array} \right.$$

$$\left\{ \begin{array}{l} \forall m, k \quad \sum_n B_n X_{mn} \geq \sum_{k'} b_{k'} U_{\min(k, k') \max(k, k') m}; \\ \forall m, k, k_1 \leq k \quad Y_{km} + Y_{k_1 m} - U_{k_1 k m} \leq 1; \\ \forall m, k, k_1 \leq k \quad -Y_{km} - Y_{k_1 m} + 2U_{k_1 k m} \leq 0. \end{array} \right.$$

$$(k = \overline{1, K}; \quad k' = \overline{1, K}; \quad m = \overline{1, M}; \quad n = \overline{1, N})$$

7.8.2 Алгоритм решения задачи псевдодобулевой оптимизации

Полученную задачу можно решать несколькими способами. Например, можно воспользоваться методом отсечений Гомори или прибегнуть к методу неявного перебора — методу ветвей и границ для решения линейной задачи с булевыми переменными [58].

Остановимся на описании одного из методов неявного перебора — метода псевдодобулевой оптимизации с помощью процедуры Дэвиса–Путнама, предложенного Питером Бартом (Институт Информатики Макса Планка, MPI) [124].

Основные понятия. Пусть B — конечный набор булевых переменных. Символ L_i — либо булева переменная X_i , либо $\bar{X}_i = 1 - X_i$. Под $Var(X_j) = Var(\bar{X}_j) = X_j$ будем подразумевать переменную, обозначаемую данным символом. *Линейное псевдодобулево выражение* — выражение вида $c_1 L_1 + \dots + c_n L_n$ (обозначается также cL), где $c_i \in \mathbf{Z}$. Пусть T — множество всех линейных псевдодобулевых выражений. *Назначение переменных* — отображение $\alpha: B \rightarrow \{0, 1\}$. Также под этим термином понимают отображение $\alpha: T \rightarrow \mathbf{Z}$.

Линейное псевдодобулево неравенство — неравенство вида $cL \geq d$. *Расширение (extension) $Ext(S)$* системы линейных псевдодобулевых неравенств S — множество назначений переменных, удовлетворяющих S . Система линейных псевдодобулевых неравенств S *мажорирует* систему S' , если $Ext(S)$ является подмножеством $Ext(S')$. Линейное псевдодобулево неравенство $cL \geq d$, в котором $c_i = 1 \forall 1 \leq i \leq n$, называется *расширенным ограничением (extended clause)* и обозначается $L \geq d$. Если дополнительно $d = 1$, то линейное псевдодобулево неравенство называется *классическим ограничением (classical clause)*. Задача о разрешимости системы ограничений — так называемая *SAT-задача (propositional satisfiability problem)*.

Линейное псевдобулево неравенство представлено в *нормальной форме*, если

$$d \geq c_1 \geq \dots \geq c_n \geq 1 \text{ и } \text{Var}(L_i) \neq \text{Var}(L_j) \forall 1 \leq i \neq j \leq n. \quad (7.44)$$

Полагаем, что $d \geq 1$, так как в противном случае псевдобулево неравенство в нормальной форме является тождеством, т.е. верно при любых назначениях переменных α , и может быть исключено из дальнейшего рассмотрения.

Теорема. Для каждого линейного псевдобулева неравенства существует ему эквивалентное в нормальной форме.

Доказательство. Пусть дано произвольное линейное псевдобулево неравенство

$$e_1 L'_1 + \dots + e_m L'_m \geq d'. \quad (7.45)$$

Группируя слагаемые, содержащие одинаковые переменные, получаем

$$a_1 X_1 + b_1 \overline{X_1} + \dots + a_n X_n + b_n \overline{X_n} \geq d,$$

где X_i попарно различны. Если $a_i = b_i$, можно упростить $a_1 X_1 + b_1 \overline{X_1}$ до a_i и перенести в правую часть. Если же $a_i \neq b_i$, заменяем $a_1 X_1 + b_1 \overline{X_1}$ на

$$c'_i L_i + c''_i := \begin{cases} (a_i - b_i) X_i + b_i, & \text{если } a_i > b_i, \\ (b_i - a_i) \overline{X_i} + a_i, & \text{если } b_i > a_i. \end{cases}$$

Заметим, что все c'_i положительны. Переносим все константы c''_i в правую часть, получим в правой части $d = d' - \sum_{i=1}^n c''_i$. Перенумеруем переменные в соответствии с требованием невозрастания коэффициентов в порядке следования. Получим неравенство

$$c'_1 L'_1 + \dots + c'_n L'_n \geq d, \quad (7.46)$$

где $c'_1 \geq \dots \geq c'_n \geq 1$, $\text{Var}(L_i) \neq \text{Var}(L_j) \forall 1 \leq i < j \leq n$, $d \geq 1$, так как в противном случае (7.45) — тождество.

Предположим, что для некоторого i $c_i \geq d$. Тогда при любом назначении переменных, таком, что $\alpha(L_i) = 1$, значение левой части (7.46) больше d . При любом назначении переменных, таком, что $\alpha(L_i) = 0$, значение левой части не зависит от c_i . Поэтому можно определить

$$c_i := \begin{cases} c'_i, & \text{если } c'_i \leq d, \\ d, & \text{если } c'_i > d. \end{cases} \quad (7.47)$$

Таким образом, получили линейное псевдобулево неравенство в нормальной форме

$$c_1 L_1 + \dots + c_n L_n \geq d, \quad (7.48)$$

эквивалентное исходному.

Ниже предполагаем, что все линейные псевдобулевы неравенства даны в нормальной форме.

Процедура Дэвиса–Путнама (DP-процедура). Единичным ограничением называется классическое ограничение $L \geq 1$, если $|L| = 1$, т.е. в левой части неравенства всего один символ. Определим шаг единичного преобразования:

$$\text{ures}(\{L_i\} \geq 1, L \geq 1) := \begin{cases} L / \{\overline{L_i}\} \geq 1, & \text{если } \overline{L_i} \in L, \\ T, & \text{если } L_i \in L, \\ L \geq 1, & \text{иначе.} \end{cases} \quad (7.49)$$

(Через T будем обозначать тождество, через $\perp$ — неравенство, множество решений которого пусто.) Для системы классических ограничений S определим

$$\begin{aligned} \text{ures}(\{L_i\} \geq 1, S) &:= \{\text{ures}(\{L_i\} \geq 1, L \geq 1) \mid L \geq 1 \in S, \\ \text{ures}(\{L_i\} \geq 1, L \geq 1) &\neq T\}, \end{aligned} \quad (7.50)$$

т.е. правило (7.49) применяется ко всем неравенствам системы, и в случае возникновения тождества ограничение из системы исключается. Фактически, мы фиксируем переменную $\text{Var}(L_i)$. Единичное преобразование (*unit resolution*) получается в результате последовательного применения правила (7.50) к системе ограничений до тех пор, пока в ней содержится единичное ограничение. Парой $\langle S, U \rangle$ обозначим SAT-задачу, к которой (возможно) были применены шаги единичного преобразования. Здесь U — набор зафиксированных переменных, S — преобразованная система ограничений.

$$\frac{\langle S, U \rangle}{\langle \text{ures}(\{L_i\} \geq 1, S), U \cup \{L_i\} \rangle}, \quad \text{если } \{L_i\} \geq 1 \in S. \quad (7.51)$$

Обозначим $\langle ur(S), ul(S) \rangle$ нормальную форму, полученную в результате применения правила (7.51) максимально возможное число раз к $\langle S, \emptyset \rangle$.

$$\text{Ext}(ur(S) \cup \{ul(S) \geq |ul(S)|\}) = \text{Ext}(S). \quad (7.52)$$

Для заданной системы классических ограничений S DP-процедура решает SAT-задачу обходом дерева, каждая вершина которого является SAT-задачей. Определим процедуры обхода дерева, оперирующие элементами $\langle P, Sol \rangle$. Здесь P — совокупность SAT-задач $\langle S, U \rangle$, Sol — совокупность назначений переменных, обращающих систему ограничений исходной SAT-задачи в систему тождеств.

$$\begin{aligned} dp_clash &: \frac{\langle \langle S, U \rangle \cup P, Sol \rangle}{\langle P, Sol \rangle}, \quad \text{если } \perp \in S; \\ dp_sol &: \frac{\langle \langle S, U \rangle \cup P, Sol \rangle}{\langle P, Sol \cup \{U\} \rangle}, \quad \text{если } S = \emptyset; \\ dp_split &: \frac{\langle \langle S, U \rangle \cup P, Sol \rangle}{\left\langle \left\{ \begin{array}{l} \langle ur(S'), ul(S') \cup U \rangle, \\ \langle ur(S''), ul(S'') \cup U \rangle \end{array} \right\} \cup P, Sol \right\rangle}, \quad \text{если } \perp \in S, S = \emptyset, \end{aligned}$$

$$L_i = \text{select_literal}(S), \quad S' = S \cup \{\{L_i\} \geq 1\}, \quad S'' = S \cup \{\{\overline{L_i}\} \geq 1\}.$$

В процедуре dp_split выбирается символ ветвления L_i с помощью процедуры select_literal . Последовательное применение правил обхода к $\langle \{ur(S), ul(S)\}, \emptyset \rangle$ дает $\langle \emptyset, Sol \rangle$. Каждый элемент $U \in Sol$ определяет решение исходной системы ограничений назначением $\alpha(L_i) = 1 \quad \forall L_i \in S$, значение остальных переменных выбирается произвольно.

Так как нас интересует лишь вопрос о разрешимости системы ограничений, процесс обхода дерева останавливается, как только найдено первое решение.

DP-процедура для линейных псевдобулевых неравенств. Обобщим DP-процедуру на случай задачи о совместности системы линейных псевдобулевых неравенств, расширив понятие единичного преобразования до псевдобулева единичного преобразования.

Лемма 1. Линейное псевдобулево неравенство $cL \geq d$ мажорирует неравенство $\{L_i\} \geq 1$ тогда и только тогда, когда $c_i L_i \in cL$ и $\sum c - c_i < d$.

Доказательство. Пусть $c'L' = cL/\{c_i L_i\}$ и α — решение неравенства $cL \geq d$. Тогда $\alpha(c_i L_i) + \alpha(c'L') \geq d$ и $\sum c' = \sum c - c_i$.

$\Leftarrow$: Пусть $\sum c' < d$. Так как $\alpha(c'L') \leq \sum c' < d$, то должно быть $\alpha(c_i L_i) > 0$, таким образом, имеем $\alpha(L_i) = 1$. Значит, $\{L_i\} \geq 1$ мажорируется неравенством $cL \geq d$.

$\Rightarrow$: Предположим, что $c' \geq d$, и покажем, что $\{L_i\} \geq 1$ не мажорируется неравенством $cL \geq d$. Пусть $\alpha(L_i) = 0$ и $\alpha(L_j) = 1 \forall L_j \in L'$, тогда $\alpha(cL) = \alpha(c_i L_i) + \alpha(c'L') = \sum c' \geq d$. Таким образом, α — решение $cL \geq d$, но не $\{L_i\} \geq 1$.

Если линейное псевдобулево неравенство $cL \geq d$ мажорирует $\{L_i\} \geq 1$, L_i называется *единичным символом (unit literal)* неравенства $cL \geq d$.

Лемма 2. Если линейное псевдобулево неравенство $cL \geq d$ мажорирует $\{L_i\} \geq 1$ и $c_i L_i \in cL$, то $cL \geq d$ мажорирует $\{L_j\} \geq 1$ для всех $c_j L_j \in cL$, для которых $c_j \geq c_i$.

Доказательство. Если $\sum c - c_i < d$, то $\sum c - c_j < d$, так как $c_j \geq c_i$. По лемме 1 получаем требуемое утверждение.

Из леммы 2 следует, что линейное псевдобулево неравенство $cL \geq d$ мажорирует $\{L_i\} \geq 1$ для некоторого L_i в том и только в том случае, если $cL \geq d$ мажорирует $\{L_1\} \geq 1$, так как при L_1 стоит наибольший коэффициент $cL \geq d$ дано в нормальной форме). Определим

$$\text{fixed}(cL \geq d) := \begin{cases} L_1, & \text{если } \sum c - c_1 < d, \\ \perp, & \text{иначе} \end{cases} \quad (7.53)$$

$\text{fixed}(cL \geq d) = L_1$ тогда и только тогда, когда $cL \geq d$ мажорирует $\{L_1\} \geq 1$. Кроме того, если $\text{fixed}(cL \geq d) = \perp$, то $cL \geq d$ не мажорирует ни одно единичное ограничение. При наличии единичного ограничения $\{L_i\} \geq 1$ линейное псевдобулево неравенство $cL \geq d$ можно упростить: L_i заменяется на 1, и константа c_i переносится в правую часть. Определим

$$\text{fix}(L_i, cL \geq d) := \begin{cases} T, & \text{если } d - c_i \leq 0 \\ & \text{и } c_i L_i \in cL, \\ cL/\{c_i L_i\} \geq d - c_i, & \text{если } d - c_i > 0 \\ & \text{и } c_i L_i \in cL, \\ \perp, & \text{если } \sum c - c_i < d \\ & \text{и } c_i \bar{L}_i \in cL, \\ cL/\{c_i \bar{L}_i\} \geq d, & \text{если } \sum c - c_i \geq d \\ & \text{и } c_i \bar{L}_i \in cL, \\ cL \geq d, & \text{если } L_i \notin L \\ & \text{и } \bar{L}_i \notin L. \end{cases} \quad (7.54)$$

Для системы линейных псевдобулевых неравенств S будем подразумевать под $\text{fix}(L_i, S)$ систему неравенств $\text{fix}(L_i, cL \geq d) \neq T, cL \geq d \in S$. Определим *псевдобулево единичное преобразование (pseudo-Boolean unit resolution)* системы

линейных псевдобулевых ограничений S :

$$\frac{\langle S, U \rangle}{\langle fix(L_i, S), U \cup \{L_i\} \rangle}, \text{ если } cL \geq d \in S \text{ и } L_i = fixed(cL \geq d) (\neq \perp). \quad (7.55)$$

Обозначим $\langle pbur(S), pbul(S) \rangle$ нормальную форму, полученную применением преобразования (7.50) к $\langle S, \emptyset \rangle$ максимально возможное число раз:

$$Ext(pbur(S) \cup \{pbul(S) \geq |pbul(S)|\}) = Ext(S). \quad (7.56)$$

Обобщим DP-процедуру на случай системы линейных псевдобулевых ограничений.

$$pbdp_clash : \frac{\langle \langle S, U \rangle \cup P, Sol \rangle}{\langle P, Sol \rangle}, \text{ если } \perp \in S;$$

$$pbdp_sol : \frac{\langle \langle S, U \rangle \cup P, Sol \rangle}{\langle P, Sol \cup \{U\} \rangle}, \text{ если } S = \emptyset;$$

$$pbdp_split : \frac{\langle \langle S, U \rangle \cup P, Sol \rangle}{\left\langle \left\{ \begin{array}{l} \langle pbur(S'), pbul(S') \cup U \rangle, \\ \langle pbur(S''), pbul(S'') \cup U \rangle \end{array} \right\} \cup P, Sol \right\rangle},$$

если $\perp \in S, S = \emptyset$;

$$L_i = select_literal(S), S' = S \cup \{\{L_i\} \geq 1\}, S'' = S \cup \{\{\bar{L}_i\} \geq 1\}.$$

Последовательное применение правил обхода к $\{\langle pbur(S), pbul(S) \rangle, \emptyset\}$ дает $\langle \emptyset, Sol \rangle$.

Оптимизация с помощью псевдобулевой DP-процедуры. Задача поиска глобального экстремума линейного псевдобулева многочлена $c'L'$ (задача оптимизации) может быть решена путем решения последовательности задач о совместности линейной псевдобулевой системы ограничений. Рассмотрим задачу поиска максимума (для минимума — аналогично).

Будем искать решение α системы S , такое, что $\alpha(c'L') \geq \alpha'(c'L')$ для любого решения α' . Рассмотрим последовательность задач о совместности системы $S_i = S \cup \{c'L' \geq \max_i\}$. $c'L' \geq \max_i$ называется *неравенством целевой функции*. Пусть значение $\max_0$ таково, что $c'L' \geq \max_0$ — тождество. Решая S_0 , получим решение α_0 , являющееся также решением S . Нижняя граница максимума целевой функции тогда $\alpha_0(c'L')$. Определим $\max_{i+1} = \alpha_i(c'L') + 1$, исключив, таким образом, назначения переменных, не дающие улучшения нижней границы максимума целевой функции по сравнению с текущей. Если S_i совместна, а S_{i+1} — нет, то $\alpha_i(c'L')$ — искомый максимум.

Используя этот прием, модифицируем pbdp-методы для решения задачи оптимизации. Для этого заменим процедуру pbdp_sol. Вместо добавления к множеству Sol вычисленного решения, фиксирующего символы U , строим решение α , такое, что $\alpha(L_i) = 1 \forall L_i \in U \cup (L' \setminus \bar{U})$. Тогда $\alpha(c'L')$ — максимальное значение $c'L'$ при условии $\{U \geq |U|\}$. Затем добавляем к остальным SAT-задачам линейное псевдобулево неравенство $c'L' \geq \alpha(c'L') + 1$. Таким образом, каждое последующее решение будет давать большее значение целевой функции по сравнению с преды-

душим. При отсутствии последующих решений последнее вычисленное значение $\alpha(c'L')$ будет искомым.

Пусть $c'L'$ — линейное псевдобулево выражение с положительными коэффициентами, ($c_i > 0 \forall c_i \in c'$). Определим процедуру обхода, оперирующую элементами $\langle P, \max \rangle$, где P — совокупность пар $\langle S, U \rangle$, S — система псевдобулевых ограничений, U — набор символов с зафиксированными значениями, $\max$ — текущее значение нижней границы максимума целевой функции:

$$opbdp_clash : \frac{\langle \langle S, U \rangle \cup P, \max \rangle}{\langle P, \max \rangle}, \text{ если } \perp \in S;$$

$$opbdp_climb : \frac{\langle \langle S, U \rangle \cup P, \max \rangle}{\langle propagate(P, \max' + 1), \max' \rangle},$$

$$S = \emptyset,$$

$$\text{если } \forall L_i \in U \cup (L' \setminus \overline{U}) : \alpha'(L_i) = 1,$$

$$\max' = \alpha'(c'L');$$

$$opbdp_split : \frac{\langle \langle S, U \rangle \cup P, \max \rangle}{\left\langle \left\{ \begin{array}{l} \langle pbur(S'), pbul(S') \cup U \rangle, \\ \langle pbur(S''), pbul(S'') \cup U \rangle \end{array} \right\} \cup P, \max \right\rangle},$$

$$\text{если } \perp \in S, \quad S = \emptyset;$$

$$L_i = select_literal(S), \quad S' = S \cup \{ \{L_i\} \geq 1 \}, \quad S'' = S \cup \{ \{ \overline{L_i} \} \geq 1 \};$$

$$propagate(\langle S, U \rangle \cup P, \max) :=$$

$$:= \{ \langle pbur(S \cup \{c'L' \geq \max\}), pbul(S \cup \{c'L' \geq \max\}) \cup U \rangle \} \cup$$

$$\cup propagate(P, \max);$$

$$propagate(\emptyset, \max) := \emptyset.$$

Последовательное применение правил обхода к $\langle \{ \langle pbur(S), pbul(S) \rangle \}, -1 \rangle$ дает $\langle \emptyset, \max \rangle$, где $\max \geq \alpha(c'L') \forall \alpha \in Ext(S)$ и $\max = -1$ тогда и только тогда, когда S несовместна.

Описанная схема вычислений была реализована программно. В файле с входными данными (input.xml) задаются координаты мест-кандидатов на размещение станций, а также параметры моделей станций: радиусы зон покрытия R_1, R_2 , стоимости подключения клиентов к станции W_1, W_2 , производительность B , стоимость C .

Для решения задачи псевдобулевой оптимизации использована программная реализация описанного алгоритма — opbdp [124]. Результатом работы программы являются матрица размещения станций (X_{mn}), матрица подключения клиентов (Y_{km}) и полная стоимость комплекса. Количество ограничений и количество переменных в задаче псевдобулевой оптимизации для условий (7.41) и (7.42) представлены в табл. 7.6.

Вместо условия (7.42) можно рассмотреть ограничение на максимальное количество клиентов R_n , обслуживаемых станцией (оно также может быть получено, если задана минимально допустимая ширина канала для клиентов в сети

и известна производительность каждой станции):

$$\forall m \sum_n R_n X_{mn} \geq \sum_k Y_{km}.$$

В этом случае количество ограничений и количество переменных будут линейно зависеть от K , как и в случае неравенств (7.41).

Таблица 7.6. Количество ограничений и количество переменных в задаче псевдо-булевой оптимизации

	(7.41)	(7.42)
Кол-во ограничений	$2(MN + 1)K + 2M$	$MK^2 + M +$ $+2(M(N + 1) + 1)K$
Кол-во переменных	$M(N + 1)K + MN$	$M(K^2 + K)/2 +$ $+M(N + 1)K + MN$

Альтернативная формулировка задачи. Имеется M мест-кандидатов на размещение станций с известными координатами, N станций, K клиентов. Для каждой станции известно ограничение на суммарную ширину канала всех подключенных к ней клиентов (производительность) B_n . Для каждого клиента заданы координаты и требуемая ширина канала b_k .

C_{mn} — стоимость установки n -й станции на m -м месте.

Стоимость подключения клиента к n -й станции:

$$\begin{cases} W_n^1 > 0, & \text{если расстояние между ними} \leq R_1; \\ W_n^2 > 0, & \text{если расстояние между ними } R_1 < r \leq R_2. \end{cases}$$

Требуется разместить станции на местах-кандидатах и подключить клиентов к станциям, минимизировав стоимость комплекса.

Введем переменные:

$$X_{mn} = \begin{cases} 1, & \text{если } n\text{-я станция размещена на } m\text{-м месте;} \\ 0, & \text{в противном случае.} \end{cases}$$

$$Y_{km} = \begin{cases} 1, & \text{если } k\text{-й клиент подключен к } n\text{-й станции;} \\ 0, & \text{в противном случае.} \end{cases}$$

Пусть также W_{mnk} — стоимость подключения k -го клиента к n -й станции при условии, что она размещена на m -м месте.

1. Каждая станция может быть размещена не более чем на одном месте-кандидате:

$$\forall n \quad \sum_m X_{mn} \leq 1.$$

2. Требуется подключить всех клиентов:

$$\forall k \quad \sum_n Y_{nk} = 1.$$

3. Ограничение на суммарную ширину каналов для клиентов, подключенных к станции:

$$\forall n \quad \sum_k b_k Y_{nk} \leq B_n.$$

4. Возможность подключения клиента к установленной станции:

$$\forall k \quad \sum_m \sum_n W_{mnk} X_{mn} Y_{nk} > 0.$$

Целевая функция (стоимости):

$$\Phi = \sum_{\substack{m,n,k \\ W_{mnk} > 0}} \sum W_{mnk} X_{mn} Y_{nk}.$$

Далее решаем задачу: $\Phi \rightarrow \min$ с ограничениями (7.37), (7.38), (7.39), (7.40); X_{mn}, Y_{nk} — булевы переменные.

Можно преобразовать задачу к виду:

$$\sum_{\substack{m,n,k \\ W_{mnk} > 0}} W_{mnk} Z_{mnk} \sum_m \sum_n C_{mn} X_{mn} \rightarrow \min$$

с ограничениями:

$$\left\{ \begin{array}{ll} \forall m, n, k & X_{mn} + Y_{nk} - Z_{mnk} \leq 1; \\ \forall m, n, k & -X_{mn} - Y_{nk} + 2Z_{mnk} \leq 0; \\ \forall n & \sum_m X_{mn} \leq 1; \\ \forall k & \sum_n Y_{nk} = 1; \\ \forall n & \sum_k b_k Y_{nk} \leq B_n; \\ \forall k & \sum_m \sum_n W_{mnk} Z_{mnk} > 0. \end{array} \right.$$

ГЛАВА 8

Стандарт широкополосного доступа IEEE 802.16-2004

При переходе к созданию систем широкополосного радиодоступа с интеграцией услуг стало понятно, что основополагающие принципы, заложенные в беспроводные системы на предыдущих этапах, нуждаются в существенной корректировке. На сигнальном уровне первостепенное значение приобрело оптимальное использование спектрального ресурса радиоканала при любых соотношениях «скорость–помехоустойчивость». На уровне протоколов стало необходимым обеспечивать заданный уровень качества обслуживания (QoS) любому абоненту сети.

Разработанный IEEE стандарт IEEE 802.16-2004 [164] представляет собой рассчитанную на внедрение в городских распределенных (региональных) беспроводных сетях (WirelessMAN) технологию беспроводного широкополосного доступа операторского класса. В последнем — его основное отличие от группы стандартов IEEE 802.11, ориентированных на работу в безлицензионном диапазоне. Следует отметить, что в Европе Институт стандартизации в области телекоммуникаций ETSI принял континентальный эквивалент стандарта IEEE 802.16, именуемый ETSI HIPERMAN. Отметим, что HIPERMAN — это подмножество IEEE 802.16, он предусматривает работу в диапазоне 2–11 ГГц и только в одном из режимов — WirelessMAN-OFDM (см. далее).

Часто используют коммерческое имя стандарта беспроводной связи IEEE 802.16 — WiMax (Worldwide Interoperability for Microwave Access). Произошло оно от названия международной организации WiMax Forum (www.wimaxforum.org), в которую входят ряд ведущих коммуникационных и полупроводниковых компаний (Airspan Networks, Alvarion Ltd, Aperto Networks, Fujitsu Microelectronics America, Intel, OFDM Forum, Proxim Corporation, Wi-LAN Inc и др.). Однако следует помнить, что на самом деле WiMax, равно как и европейский HIPERMAN, рассматривает режим WirelessMAN-OFDM.

WiMax Forum был организован 11 апреля 2003 года. Его целью является содействие разработке беспроводного оборудования для доступа к широкополосным сетям, скорейшее развертывание сетей во всем мире и сертификация оборудования IEEE 802.16, а также подготовка спецификаций, призванных обеспечить

совместимость оборудования разных производителей. Одна из целей WiMax — дальнейшее разделение труда на рынке производителей беспроводного оборудования. Поставщики элементной базы, такие как Intel и Fujitsu, будут разрабатывать ее для всех производителей оборудования, а производители оборудования смогут сконцентрировать свои усилия на оборудовании со стандартной элементной базой. По данным аналитиков, члены WiMax Forum представляют собой более 75% рынка производителей оборудования широкополосного беспроводного доступа. Ожидается, что WiMax Forum начнет сертификацию оборудования в июле 2005 года. В качестве сертификационной лаборатории уже выбрана испанская компания Cetecom. По условиям WiMax сертификация должна проводиться с проверкой совместимости оборудования трех производителей с одинаковыми профилями.

Стандарт IEEE 802.16 описывает принципы построения сетей регионального масштаба в диапазонах до 66 ГГц, точнее, их физический и MAC-уровни. Для этого стандарт предусматривает пять режимов (табл. 8.1). Сразу оговоримся, что только один из них — WirelessMAN-SC предназначен для работы в диапазоне 10–66 ГГц. Он ориентирован на магистральные сети (точка-точка, точка-многоточка), работающие в режиме прямой видимости (так как затухание столь высокочастотных сигналов при отражении очень велико), с типичными скоростями потока данных (bit stream) 120 Мбит/с и шириной канала порядка 25 МГц.

Остальные режимы разработаны для диапазонов менее 11 ГГц. Один из них — WirelessMAN-SCa — это «низкочастотная» вариация WirelessMAN-SC (с рядом дополнительных механизмов). Режим WirelessHUMAN представляет собой адаптацию остальных методов для работы в безлицензионных (в США) диапазонах. Остальные два режима — WirelessMAN-OFDM и WirelessMAN-OFDMA — это совсем новые методы (ранее входили в утвержденный в 2003 году стандарт IEEE 802.16a, но с тех пор претерпели ряд изменений).

Отметим, что все режимы диапазона ниже 11 ГГц отличает три характерные детали — механизмы автоматического запроса повторной передачи (ARQ — automatic repeat request), поддержка работы с адаптивными антенными системами (AAS — adaptive antenna system) и пространственно-временное кодирование (STC — space time coding). Кроме того, помимо централизованной архитектуры «точка-многоточка», в режиме OFDM предусмотрена поддержка архитектуры Mesh-сети («сетки» — децентрализованной сети взаимодействующих друг с другом систем). Примечательно, что если в документе IEEE 802.16a речь шла о диапазоне 2–11 ГГц, то в новом стандарте нижняя граница так четко не оговаривается (упоминается «как правило, не ниже 1 ГГц»). Хотя по информации ряда аналитиков известно, что всерьез анализируется возможность использования диапазона 700 МГц для модификации стандарта для мобильных абонентов IEEE 802.16e.

Еще одна особенность стандарта — в нем предусмотрен режим работы в безлицензионном (в США) диапазоне WirelessHUMAN (High-speed Unlicensed Metropolitan Area Network). Фактически речь идет об адаптации описанных методов в полосе частот порядка 5–6 ГГц. Основные отличия WirelessHUMAN — это использование только временного дуплексирования, режим динамического распределения частот (DFS — dynamic frequency selection) и механизм сквозной

нумерации частотных каналов. Однако, поскольку в России безлицензионных диапазонов в гигагерцовой области нет и ничего подобного нам не грозит, подробно останавливаться на данном режиме не будем.

Таблица 8.1. Основные режимы в стандарте IEEE 802.16-2004

Режим	Частотный диапазон	Опции	Метод дуплексирования
WirelessMAN-SC	10–66 ГГц		TDD/FDD
WirelessMAN-SCa	< 11 ГГц	AAS/ARQ/STC/256-QAM	TDD/FDD
WirelessMAN-OFDM	< 11 ГГц	AAS/ARQ/STC/Mesh	TDD/FDD
WirelessMAN-OFDMA	< 11 ГГц	AAS/ARQ/STC	TDD/FDD
WirelessHUMAN	< 11 ГГц*	DFS/AAS/ARQ/Mesh/STC	TDD

*Безлицензионный диапазон (в США и Европе).

8.1 МАС-уровень стандарта IEEE 802.16

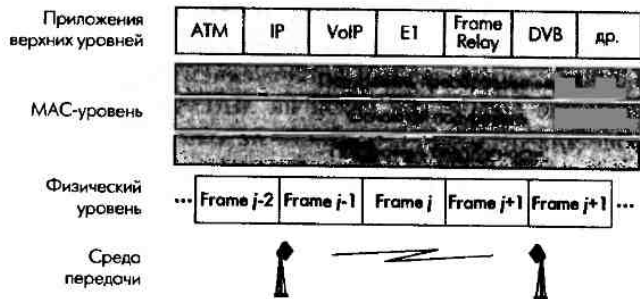
8.1.1 Структура МАС-уровня

Задачи непосредственной доставки потоков данных между БС и абонентскими станциями решаются на физическом уровне стандарта IEEE 802.16. Функции же, связанные с формированием структур этих данных, а также управлением работой системы IEEE 802.16, реализуются на МАС-уровне. Оборудование стандарта IEEE 802.16 призвано формировать транспортную среду для различных приложений (сервисов), поэтому первая задача, решаемая в IEEE 802.16, это механизм поддержки разнообразных сервисов верхнего уровня. Разработчики стандарта стремились создать единый для всех приложений протокол МАС-уровня, независимо от особенностей физического канала. Это существенно упрощает связь терминалов конечных пользователей с городской сетью передачи данных — физической среды передачи в разных фрагментах WMAN могут быть различны, но структура данных одина. В одном канале могут работать (не одновременно) сотни различных терминалов еще большего числа конечных пользователей. Этим пользователям необходимы самые разные сервисы (приложения) — потоки голоса и данных с временным разделением, соединения по протоколу IP, пакетная передача речи через IP (VoIP) и т. п. Более того, качество услуг (QoS) каждого отдельного сервиса не должно изменяться при работе через сети IEEE 802.16. Алгоритмы и механизмы доступа МАС-уровня должны уверенно решать все эти задачи.

Структурно МАС-уровень стандарта IEEE 802.16 подразделяется на три подуровня (рис. 8.1) — подуровень преобразования сервиса CS (Convergence Sublayer), основной подуровень CPS (Common Part Sublayer) и подуровень защиты PS

(Privacy Sublayer). На подуровне защиты реализуются функции криптозащиты данных и механизмы аутентификации/предотвращения несанкционированного доступа. Для этого предусмотрены два основных компонента — набор алгоритмов криптозащиты и протокол управления ключом шифрования. Ключ каждой абонентской станции (АС) базовая станция (БС) может передавать в процессе авторизации, используя схему работы «клиент (АС) — сервер (БС)».

Рис. 8.1. MAC-уровень в стандарте IEEE 802.16



На подуровне преобразования сервиса происходит трансформация потоков данных протоколов верхних уровней для передачи через сети IEEE 802.16. Для каждого типа приложений верхних уровней стандарт предусматривает свой механизм преобразования. На сегодня описаны и вошли в спецификацию IEEE 802.16 два основных типа сервисных потоков — ATM и пакетная передача. Последняя подразумевает достаточно широкий набор различных протоколов — IP, Ethernet (IEEE 802.3), виртуальные ЛВС (VLAN, IEEE 802.1Q-1998). Цель работы на CS-подуровне — оптимизация передаваемых потоков данных каждого приложения верхнего уровня с учетом их специфики. Поэтому важнейшая задача, решаемая на данном подуровне, — задача классификации пакетов/ячеек. От ее результатов зависит и оптимизация передаваемых потоков, и выделение полосы пропускания для каждого из них.

Для оптимизации транслируемых потоков предусмотрен специальный механизм удаления повторяющихся фрагментов заголовков PHS (Payload Header Suppression). Действительно, и в ATM, и в пакетном режиме данные передаются отдельными порциями — ячейками и пакетами. Каждая такая порция данных состоит, в общем случае, из заголовка и поля данных — фиксированных размеров для ячеек ATM (5 и 48 байт соответственно) и достаточно произвольных при пакетной передаче. Во многих случаях заголовки пакетов и ячеек содержат повторяющуюся информацию, излишнюю при трансляции посредством протокола IEEE 802.16. Механизм PHS позволяет избавиться от передачи избыточной информации: на передающем конце пакеты приложений в соответствии с определенными правилами преобразуются в структуры данных MAC-уровня IEEE 802.16, на приемном — восстанавливаются.

8.1.2 Соединения и сервисные потоки

Ключевой момент в стандарте IEEE 802.16 — это понятие сервисного потока и связанные с ним понятия «соединение» и «идентификатор соединения» (CID).

Поскольку система IEEE 802.16 лишь транспортная среда, ее инфраструктура фактически формирует коммуникационные каналы для потоков данных различных приложений верхних уровней (сервисов) — передача видеоданных, ATM-потоки, IP-потоки, передача телефонных мультиплексированных пакетов типа E1 и т.д. Каждое из таких приложений обладает своими требованиями к скорости передачи, надежности (качеству обслуживания, QoS), криптозащите и т.д. Соответственно, и данные каждого приложения следует передавать через транспортную среду с учетом этой специфики.

Сервисным потоком в стандарте IEEE 802.16 называется поток данных, связанный с определенным приложением. В этом контексте соединение — это установление логической связи на MAC-уровнях на передающей и приемной стороне для передачи сервисного потока. Каждому соединению присваивается 16-разрядный идентификатор CID, с которым однозначно связаны тип и характеристики соединения. В частности, по запросу предоставления/изменения полосы пропускания со стороны АС базовая станция сразу понимает, с каким сервисным потоком имеет дело и какие условия передачи ему нужно обеспечить. Так, при начальной инициализации в сети каждой АС назначается три CID для служебных сообщений трех уровней. Принципиально, что одна АС может устанавливать множество различных соединений с различными CID. Характерный пример, когда связь крупного офиса с телекоммуникационным узлом организована через систему IEEE 802.16. В этом случае одна АС в офисе может поддерживать совершенно разные приложения — телефонию, телевидение, доступ в Интернет и в распределенную корпоративную сеть и т.д. Каждое из этих приложений предъявляет свои требования к QoS и скорости передачи, которые нужно удовлетворить. Посредством CID базовая станция узнает, с чем имеет дело, и предоставляет необходимый ресурс.

Сервисный поток характеризуется набором требований к каналу передачи информации — времени задержки символов, уровню флуктуаций задержек (джиттеру) и гарантированной пропускной способности. Каждому сервисному потоку в сети присваивается идентификатор SFID (32 разряда), основываясь на котором БС (а в ряде случаев и АС) определяют необходимые параметры связанного с данным сервисным потоком конкретного соединения. Для общей стандартизации работы в сети используется понятие сервисного класса — устойчивого набора параметров для стандартных приложений, например для трансляции телефонных каналов E1. Параметры сервисного потока можно идентифицировать, просто указав его принадлежность к определенному сервисному классу.

8.1.3 Пакеты MAC-уровня

Весь поток данных в сетях IEEE 802.16 — это поток пакетов. На основном подуровне MAC формируются пакеты данных (MAC PDU — MAC Protocol Data Unit, блоки данных MAC-уровня), которые затем передаются на физический уровень, инкапсулируются в физические пакеты и транслируются через канал связи. Пакет MAC PDU (далее — PDU) включает заголовок и поле данных (его может и не быть), за которым может следовать контрольная сумма CRC (рис. 8.2). Заголовок PDU занимает 6 байт и может быть двух типов — общий и заголовок

запроса полосы пропускания. Общий заголовок используется в пакетах, у которых присутствует поле данных. В общем заголовке указывается идентификатор соединения (CID), тип и контрольная сумма заголовка, а также приводится информация о поле данных (см. табл. 8.2).

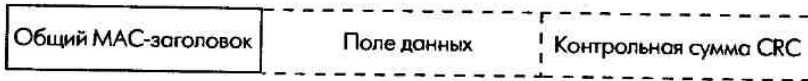


Рис. 8.2. Пакет MAC-уровня IEEE 802.16

Таблица 8.2. Структура общего заголовка MAC PDU
(от старшего к младшим битам)

Поле	Длина, бит
тип заголовка = 0	1
признак шифрования поля данных	1
тип подзаголовков	6
не используется	1
признак наличия CRC	1
индекс ключа шифрования	2
не используется	1
длина пакета, включая заголовок (байт)	11
идентификатор соединения CID	16
контрольная сумма заголовка (задающий полином $g(D) = D^8 + D^2 + D + 1$)	8

Заголовок запроса полосы (также 6 байт) применяется, когда АС просит у БС выделить или увеличить ей полосу пропускания в нисходящем канале. При этом в заголовке указывается CID и размер требуемой полосы (в байтах, без учета заголовков физических пакетов). Поля данных после заголовков запроса полосы нет.

Поле данных может содержать:

- подзаголовки MAC,
- управляющие сообщения и
- собственно данные приложений верхних уровней, преобразованные на CS-подуровне.

MAC-подзаголовки могут быть пяти типов — упаковки, фрагментации, управления предоставлением канала, а также подзаголовки Mesh-сети и подзаголовки канала быстрой обратной связи Fast Feedback.

Подзаголовок упаковки используются, если в поле данных одного PDU содержатся несколько пакетов верхних уровней; *подзаголовок фрагментирования* — если, напротив, один пакет верхнего уровня разбит на несколько PDU.

Подзаголовок управления предоставлением доступа предназначен для того, чтобы АС сообщала БС изменение своих потребностей в полосе пропускания (число байт в восходящем канале для определенного соединения, сообщение о переполнении выходной очереди в АС, требование регулярного опроса со стороны БС для выяснения необходимой полосы).

Подзаголовок Fast Feedback служит для назначения конкретной АС интервала для быстрого ответа на запрос БС (например, в ходе процедур измерения характеристик канала).

Управляющие сообщения — это основной механизм управления системой IEEE 802.16. Всего зарезервировано 256 типов управляющих сообщений, из них используется 48. Формат управляющих сообщений прост — поле типа сообщения (1 байт) и поле данных (параметров) произвольной длины.

Описание профилей пакетов, управление доступом, механизмы криптозащиты, динамическое изменение работы системы и т.д. — все функции управления, предоставления доступа, запроса и подтверждения реализуются через управляющие сообщения. Рассмотренные ниже карты входящего/нисходящего каналов (UL-/DL-MAP) также являются управляющими сообщениями.

8.1.4 Общая структура кадров IEEE 802.16

Для понимания принципов управления соединениями и потоками данных в сетях IEEE 802.16 рассмотрим общие принципы передачи в физическом канале.

Передача данных на физическом уровне происходит посредством непрерывной последовательности кадров фиксированной длительности. Кадр состоит из двух субкадров — для нисходящего потока (от базовой станции к АС) и для восходящего (от АС к БС). Дуплексный механизм предусматривает как частотное (FDD — frequency division duplex), так и временное (TDD — time division duplex) разделение восходящего и нисходящего субкадров.

При временном дуплексировании каналов кадры передаются в одном частотном диапазоне, сначала нисходящий (DL), затем — восходящий (UL). При частотном дуплексировании восходящий и нисходящий субкадры транслируются в своем частотном диапазоне с определенным разносом.

DL-субкадр начинается с синхронизирующей последовательности (преамбулы), за которой следует управляющая секция с набором широковещательных (предназначенных всем) служебных сообщений (например, дескрипторы нисходящего и восходящего каналов), включая карты нисходящего и восходящего каналов (DL-MAP и UL-MAP). За ними передаются пакеты физического уровня, содержащие как служебные сообщения, так и данные для различных АС.

Пакеты в нисходящем субкадре следуют друг за другом без интервалов. Чтобы абонентские станции могли отличить один пакет от другого, в управляющей секции передаются карты нисходящего (DL-MAP) и восходящего (UL-MAP) каналов. В карте нисходящего канала указана длительность и номер кадра, идентификационный номер базовой станции, номер последнего переданного дескриптора нисходящего канала (в котором описаны профили пакетов текущего кадра), а также точка начала и тип профиля каждого пакета. Точка начала отсчитывается в так называемых физических слотах, каждый физический слот равен четырем модуляционным символам.

Профиль пакета — это список его параметров, включая метод модуляции, тип FEC-кодирования (с параметрами схем кодирования), а также диапазон значения отношения сигнал/шум в приемном канале конкретной станции, при котором данный профиль может применяться. Список профилей в виде специальных

управляющих сообщений — дескрипторов нисходящего и восходящего каналов (DCD/UCD) — транслируется базовой станцией с периодом не менее 10 с, причем каждому профилю присваивается номер, который и используется в карте нисходящего канала.

В восходящем субкадре для каждой передающей АС базовая станция резервирует специальные временные интервалы (или частотно-временные) — тайм-слоты. Информация о распределении тайм-слотов между АС записывается в карте восходящего канала UL-MAP, транслируемого в каждом кадре. UL-MAP функционально аналогична DL-MAP — в ней сообщается, сколько тайм-слотов в субкадре, точка начала и идентификатор соединения для каждого из них, а также типы профилей всех пакетов. Сообщение UL-MAP текущего кадра может относиться как к данному кадру, так и к последующему.

Кроме назначенных БС тайм-слотов для определенных АС, в UL-канале предусмотрены интервалы конкурентного доступа, в течение которых АС может передать сообщение для первичной регистрации в сети или для запроса канала/изменения полосы пропускания предоставленного канала.

8.1.5 Принцип предоставления канальных ресурсов

Основной принцип предоставления доступа к каналу в стандарте IEEE 802.16 — это доступ по запросу Demand Assigned Multiple Access (DAMA). Ни одна АС не может ничего передавать, кроме запросов на регистрацию и предоставление канала, пока БС не разрешит ей этого, т. е. отведет временной интервал в восходящем канале и укажет его расположение в карте UL-MAP. Абонентская станция может как запрашивать определенный размер полосы в канале, так и просить об изменении уже предоставленного ей канального ресурса.

Стандарт IEEE 802.16 предусматривает два режима предоставления доступа — для каждого отдельного соединения и для всех соединений определенной АС. Очевидно, что первый механизм обеспечивает большую гибкость, однако второй существенно сокращает объем служебных сообщений и требует меньшей производительности от аппаратуры.

Запросы полосы могут быть как спорадическими для БС, так и планируемыми. В первом случае запросы реализуются посредством пакетов, состоящих из заголовка запроса, передаваемых на конкурентной основе в специально выделенном для них интервале восходящего канала. Поскольку эти запросы спонтанны, в данных интервалах возможны коллизии, вызванные одновременной работой передатчиков двух и более АС.

Принцип борьбы с коллизиями аналогичен используемому в стандарте IEEE 802.11 — после того, как АС решила, что ей нужно зарегистрироваться/запросить канал, она не начинает трансляцию в первом же предназначенном для этого интервале. В АС есть генератор случайных чисел (ГСЧ), выбирающий значения из некоего диапазона от 0 до $2^n - 1$. Так, если $n = 4$, ГСЧ выбирает числа в диапазоне 0–15, например 11. Далее АС отсчитывает 11 интервалов, предназначенных для регистрации/запроса канала и только в 12-м выходит в эфир. Если передача прошла успешно и БС приняла запрос, она в определенный период ответит специальным сообщением. В противном случае АС считает попытку неудачной и

повторяет процедуру, только интервал выбора для ГСЧ удваивается. Такая последовательность действий продолжается до тех пор, пока не будет получен ответ от БС. Максимальный размер диапазона возможных значений ГСЧ ограничен — при его достижении он вновь принимает минимальное значение.

Процедура плановых запросов полосы в восходящем канале называется опросом (polling) — БС как бы запрашивает у АС их потребности в канальных ресурсах. Реально это означает, что базовая станция предоставляет конкретной АС интервал для передачи запроса о предоставлении/изменении полосы, т. е. никакой конкуренции уже нет.

Опрос может быть в «реальном времени» — интервалы для запроса предоставляются АС с тем же периодом, с каким у нее может возникнуть потребность в изменении условий доступа (например, в каждом кадре). Этот режим удобен для приложений, когда пакеты данных следуют с фиксированным периодом, но их размер не стабилен (например, видео-MPEG). Другой вариант опроса — вне «реального времени». В этом случае БС предоставляет АС интервал для запроса также периодически, но период этот существенно больше, например 1 с. Характерное приложение, для которого эффективен этот механизм, — FTP-протокол.

Для приложений, у которых периодичность и размер пакетов фиксированы (например, в телефонии шина E1), предусмотрен механизм доступа к каналу без требования (Unsolicited Grant Service — UGS). В этом случае БС предоставляет АС для передачи данных интервалы фиксированного размера с заданным периодом, соответствующим скорости потока данных.

Если в ходе работы АС нужно изменить условия доступа, она делает это посредством специального MAC-подзаголовка управления предоставлением канала. В этом подзаголовке есть специальный флаг «опроси меня», установив который, АС просит у БС интервал для запроса новой полосы. Кроме того, в этом подзаголовке АС может непосредственно запросить у БС дополнительную прибавку (в байтах) к уже предоставленной полосе для конкретного соединения (так называемый PiggyBack Request). В подзаголовке предоставления канала также есть специальный бит индикации переполнения входного буфера передатчика АС, что приводит к потере данных (slip). БС может отреагировать на появление этого сигнала, например увеличив полосу для данной АС.

8.1.6 Механизмы подтверждение приема и быстрой обратной связи

Пожалуй, самая существенная особенность MAC-уровня для диапазона ниже 11 ГГц — это процедура подтверждения приема пакетов и их повторной отправки (ARQ). Данный механизм достаточно хорошо известен и основан на способности корректирующих кодов обнаруживать ошибки передачи. Коды эти описаны в четвертой главе, поэтому лишь напомним суть механизма. Каждый пакет при передаче кодируется линейным циклическим корректирующим кодом с достаточным числом проверочных символов. На приемной стороне проверяется принадлежность принятого слова к корректирующему коду путем формирования синдрома (синдром — вектор, равный произведению вектора принятого сигнала и присущей коду матрицы проверочных уравнений). Из теории кодирования из-

вестно, что если синдром нулевой — слово кодовое, если ненулевой — некодовое. Также известно, что даже в очень плохом канале с вероятностью ошибки, стремящейся к $1/2$, вероятность ошибки декодирования можно сделать сколь угодно малой, выбирая достаточное число проверочных символов, правда, ценой существенного снижения скорости передачи.

Получив пакет, приемник обязан отправить передатчику квитанцию подтверждения (значение синдрома, ноль или нет). Все пакеты нумеруются, поэтому приемник сразу обнаруживает сбой в получении определенного пакета и сообщает об этом передатчику, который повторно отправляет недошедший пакет. Режим ARQ — опциональный и назначается конкретному соединению. Причем в рамках одного соединения весь трафик может передаваться либо с применением ARQ, либо без него, смешение недопустимо. Очевидно, что режим ARQ требует эффективного быстрого обратного канала для подтверждения. Для этого предназначено специальное сообщение. О его наличии в пакете сигнализирует признак в основном MAC-заголовке пакета.

Другая важная особенность режимов в диапазоне ниже 11 ГГц — быстрая организация канала обратной связи. Ряд механизмов стандарта, прежде всего периодическое измерение параметров канала, запросы физических параметров каналов, коррекция мощности передатчиков и др., требуют быстрого и гарантированного ответа со стороны АС. Особое значение эти процедуры приобретают при работе с адаптивными многолучевыми антенными системами, где важно определить физические параметры канала связи в каждом луче (секторе). Данные процедуры заключаются в отправке специальных сообщений и получении данных об условиях их прохождения через канал. В разных режимах (SCa, OFDM, OFDMA) названия и тип подобных сообщений несколько различны, однако сущность от этого не меняется.

В частности, для приоритетной передачи таких данных абонентскими станциями в режиме OFDMA предусмотрен специальный тип сообщений быстрой обратной связи FAST FEEDBACK и выделяется специальная область (канал) в восходящем субкадре для их передачи. Сами сообщения обратной связи представляют собой 4-разрядные двоичные числа, расположение которых в канале обратной связи для каждой АС базовая станция задает в специальном подзаголовке FAST FEEDBACK. В этом же подзаголовке указывается тип (назначение) сообщения — измерение мощности в нисходящем канале, коэффициенты антенн БС в режиме AAS и др. Предусмотрены и специальные меры для успешной доставки этих 4-разрядных сообщений. Так, каждое сообщение (4 бита) занимает один OFDMA-слот (3 символа $\times$ 1 субканал, 6 фрагментов, 48 информационных несущих). Каждой 4-битовой последовательности поставлен в соответствие фиксированный набор значений всех 48 информационных несущих в слоте. Модуляция в FAST FEEDBACK-канале — QPSK, 2 бита на символ, т. е. на каждый информационный бит сообщения приходится 24 транслируемых бита. Данная избыточность в определенной мере гарантирует достоверный прием сообщения.

8.2 Физический уровень стандарта IEEE 802.16

На физическом уровне стандарт IEEE 802.16 предусматривает три принципиально различных метода передачи данных — метод модуляции одной несущей (SC,

в диапазоне ниже 11 ГГц — SCa), метод модуляции посредством ортогональных несущих OFDM (orthogonal frequency division multiplexing) и метод мультиплексирования (множественного доступа) посредством ортогональных несущих OFDMA (orthogonal frequency division multiple access) (см. табл. 8.1). Отметим, что метод SCa отличается от своего более высокочастотного собрата SC прежде всего по методам помехоустойчивого кодирования и модуляции (допускается квадратурная 256-QAM).

8.2.1 Режим WirelessMAN-SC

Метод WirelessMAN-SC стандарта IEEE 802.16 описывает работу в диапазоне 10–66 ГГц сетей с архитектурой «точка-многоточка» (из центра — многим). Это двунаправленная система, т. е. предусмотрены нисходящий и восходящий потоки. При этом каналы подразумеваются широкополосные (до 25–28 МГц), а скорости передачи — высокие (например, 120 Мбит/с).

Канальное кодирование

Тракт обработки данных и формирования выходного сигнала для передачи через радиоканал в стандарте IEEE 802.16 достаточно обычен для современных телекоммуникационных протоколов (рис. 8.3) и практически одинаков для восходящих и нисходящих соединений.



Рис. 8.3. Тракт формирования выходного сигнала в стандарте IEEE 802.16 (нисходящий канал)

Входной поток данных скремблируется — подвергается рандомизации, т. е. умножению на псевдослучайную последовательность (ПСП), получаемую в 15-разрядном сдвиговом регистре (рис. 8.4) (задающий полином ПСП — $c(x) = x^{15} + x^{14} + 1$, начальное значение — 4A80₁₆). Далее скремблированные данные защищают посредством помехоустойчивых кодов (FEC-кодирование). При этом можно использовать одну из четырех схем кодирования: код Рида-Соломона с символами из поля Галуа GF(256), каскадный код с внешним кодом Рида-Соломона и внутренним сверточным кодом с кодовым ограничением $K = 7$ (скорость кодирования — 2/3) с декодированием по алгоритму Витерби, каскадный код с внешним кодом Рида-Соломона и внутренним кодом с проверкой на четность (8, 6, 2) и блочный турбокод. Размер кодируемого информационного блока и число избыточных байт не фиксированы — эти параметры можно задавать в

зависимости от условий среды передачи и требований QoS. Так, для кода Рида-Соломона размер исходного блока данных может быть от 6 до 255 байт, а число избыточных байт — до 32 (всего до 255 байт). Первые два алгоритма кодирования обязательны для всех устройств стандарта, остальные два алгоритма — опциональны.

Метод WirelessMAN-SC, как и следует из его названия (SC — Single Carrier), предусматривает схему с модуляцией одной несущей в каждом частотном канале. Допускается три типа квадратурной амплитудной модуляции: 4-позиционная QPSK и 16-позиционная 16-QAM (обязательны для всех устройств), а также 64-QAM (опционально). Кодированные блоки дан-

ных преобразуются в модуляционные символы (каждые 2/4/6 бит определяют один символ QPSK/16-QAM/64-QAM) в соответствии с приведенными в стандарте таблицами — каждой группе из 2/4/6 бит ставится в соответствие синфазная (I) и квадратурная (Q) координаты. Далее последовательность дискретных значений в каналах I и Q преобразуется посредством так называемого синусквадратного фильтра (square-root raised cosine filter) в непрерывные (сглаженные) сигналы. Передаточная функция синусквадратного фильтра записывается как

$$H(f) = \begin{cases} 1, & |f| < f_N(1 - \alpha) \\ \sqrt{\frac{1}{2} + \frac{1}{2} \sin \left[\frac{\pi}{2f_N} \left(\frac{f_N - |f|}{\alpha} \right) \right]}, & f_N(1 - \alpha) \leq |f| \leq f_N(1 + \alpha) \\ 0, & |f| > f_N(1 + \alpha), \end{cases}$$

где α — коэффициент избирательности (по стандарту IEEE 802.16 $\alpha = 0,25$), f_N — частота Найквиста, равная половине частоты дискретизации.

Фильтрованные потоки $I(t)$ и $Q(t)$ поступают непосредственно в квадратурный модулятор, где формируется выходной сигнал

$$S(t) = I(t) \cos(2\pi f_c t) - Q(t) \sin(2\pi f_c t),$$

f_c — несущая частота. Далее сигнал усиливается и передается в эфир. На приемной стороне все происходит в обратном порядке. В результате в зависимости от ширины канала и метода модуляции формируется достаточно широкий набор скоростей потока данных (табл. 8.3).

Структура кадров

Передача данных на физическом уровне происходит посредством непрерывной последовательности кадров. Каждый кадр имеет фиксированную длительность — 0,5; 1 и 2 мс, поэтому его информационная емкость зависит от символьной скорости и метода модуляции. Кадр состоит из преамбулы (синхропоследовательности



Рис. 8.4. Генерация ПСП для рандомизации данных

длиной 32 QPSK-символа), управляющей секции и последовательности пакетов с данными (рис. 8.5). Управляющая секция может содержать карты DL/UL-MAP и дескрипторы нисходящего/восходящего каналов. Сообщения этой секции всегда передаются посредством QPSK.

Таблица 8.3. Скорость физического потока данных в зависимости от вида модуляции и ширины канала

Ширина канала, МГц	Скорость символов, Мбод	Скорость физического потока данных, Мбит/с		
		QPSK	16-QAM	64-QAM
20	16	32	64	96
25	20	40	80	120
28	22,4	44,8	89,6	134,4

Поскольку определяемая стандартом IEEE 802.16 система двунаправленная, необходим дуплексный механизм. Он предусматривает как частотное (FDD – frequency division duplex), так и временное (TDD – time division duplex) разделение восходящего и нисходящего каналов.



Рис. 8.5. Структура кадра в стандарте IEEE 802.16 для систем с временным (а) и частотным (б) дуплексированием каналов

При временном дуплексировании каналов кадр делится на нисходящий и восходящий субкадры (их соотношение в кадре может гибко изменяться в процессе работы в зависимости от необходимой полосы пропускания для нисходящих и восходящих каналов), разделенные специальным интервалом (рис. 8.5, а). При частотном дуплексировании восходящий и нисходящий каналы транслируются каждый на своей несущей (рис. 8.5, б).

В нисходящем канале информация от базовой станции передается в виде последовательности пакетов (метод временного мультиплексирования TDM – time division multiplex) (рис. 8.6). Для каждого пакета можно задавать метод модуляции и схему кодирования данных, т. е. выбирать между скоростью и надежностью передачи. Данные о параметрах пакета, его длине, моменте начала передачи, а также о его принадлежности к определенному соединению (соответственно, об адресации определенной АС) содержатся в карте нисходящего канала DL-MAP.

Точка начала отсчитывается в так называемых физических слотах (один физический слот равен четырем модуляционным символам).

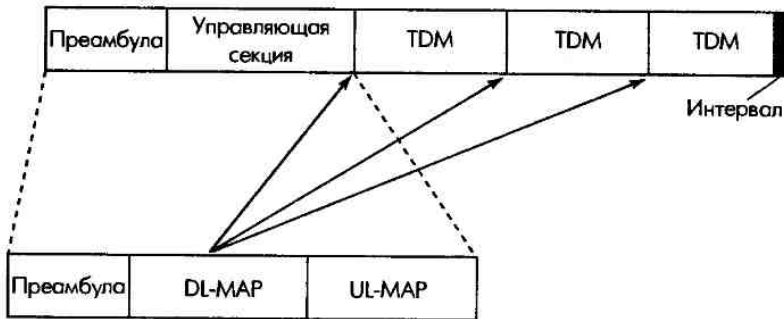


Рис. 8.6. Структура нисходящего канала

В нисходящем субкадре пакеты выстраиваются в очередь с неравной защитой пакетов так, что самые помехозащищенные передаются первыми (управляющая секция всегда передается посредством QPSK-модуляции). Если этого не сделать, абонентские станции с плохими условиями приема, которым предназначаются наиболее защищенные пакеты, могут потерять синхронизацию в ожидании своей порции информации. Пакеты в нисходящем субкадре следуют друг за другом без интервалов и предваряющих их заголовков и идентифицируются абонентскими станциями на основе информации в DL-MAP.

Абонентские станции получают доступ к среде передачи посредством механизма временного разделения каналов (TDMA — time division multiple access) (рис. 8.7). Для этого в восходящем субкадре для каждой передающей АС базовая станция резервирует специальные временные интервалы — тайм-слоты. Информация о распределении тайм-слотов между АС записывается в карте восходящего канала UL-MAP, транслируемого в каждом кадре. UL-MAP функционально аналогична DL-MAP — в ней сообщается, сколько слотов в субкадре, точка начала и идентификатор соединения для каждого из них, а также типы профилей всех пакетов. Скорость модуляции (частота символов) в восходящем канале должна быть такой же, как и в нисходящем. Отметим, что в отличие от нисходящих TDM-пакетов каждый пакет в восходящем канале начинается с преамбулы — синхропоследовательности длиной 16 или 32 QPSK-символа.

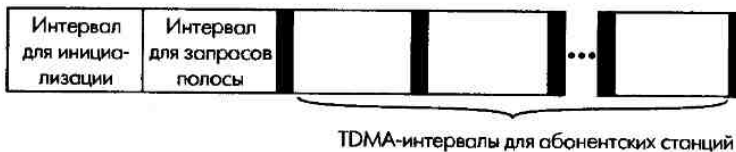


Рис. 8.7. Структура восходящего канала

В восходящем канале также предусмотрены интервалы конкурентного доступа для первичной регистрации в сети или для запроса канала/изменения полосы пропускания канала.

Стандарт IEEE 802.16 допускает применение как дуплексных, так и полудуплексных абонентских станций. Последние не способны одновременно принимать и передавать информацию. В режиме FDD для полудуплексных АС, которые в силу конструктивных особенностей сначала принимают информацию и лишь затем передают свои данные, в нисходящем FDD-кадре предусмотрена TDMA-область — для таких станций информация передается в определенных временных интервалах (рис. 8.8). Причем нисходящие пакеты, передаваемые в режиме TDMA, обязательно снабжают преамбулой — синхропоследовательностью длиной 16 QPSK-символов, чтобы полудуплексные абонентские станции могли при необходимости восстановить синхронность.

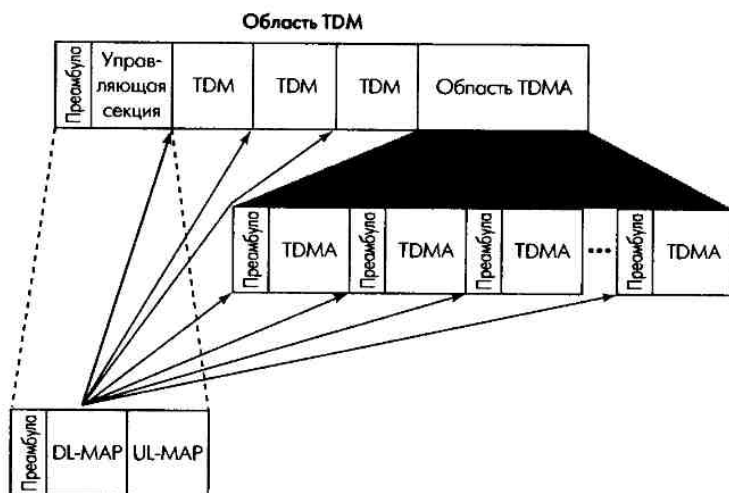


Рис. 8.8. Нисходящий канал в случае FDD при работе с полудуплексными абонентскими станциями

Отметим, что используемое в документе IEEE 802.16 терминологическое деление на TDM- и TDMA-области в нисходящем субкадре не совсем корректно, поскольку по смыслу в данном случае они практически идентичны. Видимо, проблема проистекает из первоначальной редакции стандарта IEEE 802.16-2001, в которой подразумевалось, что в режиме TDM каждая АС принимает все пакеты, декодирует их заголовки и по MAC-адресам распознает «свои». В новой редакции указано, что распределение пакетов в DL-субкадре приведено в DL-MAP (хотя не исключаются и «старая» схема). Единственное различие между TDM и TDMA — если при TDM пакеты следуют один за другим, без интервалов, то в зоне TDMA между отдельными пакетами с преамбулами возможны незаполненные интервалы, в течение которых передатчик БС «молчит».

8.2.2 Режим WirelessMAN-OFDM

Режим OFDM — это метод модуляции потока данных в одном частотном канале (шириной 1–2 МГц и более) с центральной частотой f_c (подробнее см. главу 5). Деление же на каналы, как и в случае SC, — частотное. Напомним, что при

модуляции данных посредством ортогональных несущих в частотном канале выделяются N поднесущих так, что $f_k = f_c + k\Delta f$, k — целое число из диапазона $[-N/2, N/2]$ (в данном случае $k \neq 0$). Расстояние между ортогональными несущими $\Delta f = 1/T_b$, где T_b — длительность передачи данных.

Кроме данных, в OFDM-символе передается защитный интервал длительностью T_g , так что общая длительность OFDM-символа $T_s = T_b + T_g$ (рис. 8.9). Защитный интервал представляет собой копию оконечного фрагмента символа. Его длительность T_g может составлять 1/4, 1/8, 1/16 и 1/32 от T_b .

Как указывалось в главе 5, модуляция OFDM основана на двух основных принципах: разбиение одного канала с переменными параметрами на параллельные гауссовы каналы с различными отношениями сигнал/шум и точное измерение характеристик канала.

В соответствии с первым принципом OFDM каждая поднесущая модулируется независимо посредством квадратурной амплитудной модуляции. Общий сигнал вычисляется посредством обратного быстрого преобразования Фурье (ОБПФ) как

$$s(t) = \operatorname{Re} \left\{ e^{i2\pi f_c t} \sum_{k=-N/2}^{N/2} c_k e^{i2\pi k \Delta f (t-T_g)} \right\} \quad (0 < t < T_s),$$

где c_k — комплексное представление символа квадратурной модуляции (QAM-символа). Комплексное представление удобно, поскольку генерация радиосигнала происходит посредством квадратурного модулятора в соответствии с выражением $s_k(t) = I_k \cos(2\pi f_c t) - Q_k \sin(2\pi f_c t)$, где I_k и Q_k — синфазное и квадратурное (целое и мнимое) значения комплексного символа соответственно.

Для работы алгоритмов БПФ/ОБПФ удобно, чтобы число точек соответствовало 2^m . Поэтому число несущих выбирают равным минимальному числу $N_{FFT} = 2^m$, превосходящему N . В режиме OFDM стандарта IEEE 802.16 $N = 200$, соответственно $N_{FFT} = 256$. Из них 55 ($k = -128 \dots -101$ и $101 \dots 127$) образуют защитный интервал на границах частотного диапазона канала. Центральная частота канала ($k = 0$) и частоты защитных интервалов не используются (т.е. амплитуды соответствующих им сигналов равны нулю). Оставшиеся 200 несущих — информационные.

В соответствии со вторым принципом OFDM для точного определения параметров канала необходимы так называемые пилотные несущие частоты, метод модуляции и передаваемый сигнал в которых хорошо известен всем станциям в сети. В методе OFDM предусмотрено использование восьми пилотных частот (с индексами $\pm 88, \pm 63, \pm 38, \pm 13$). Остальные 192 несущие разбиты на 16 подканалов по 12 несущих в каждом, причем в одном подканале частоты расположены не подряд. Например, подканал 1 составляют несущие с индексами $-100, -99, -98, -37, -36, -35, 1, 2, 3, 64, 65, 66$. Деление на подканалы необходимо, поскольку в

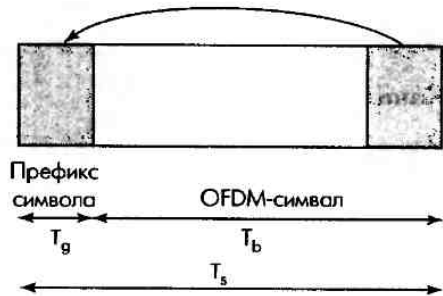


Рис. 8.9. OFDM-символ

режиме WirelessMAN-OFDM предусмотрена (опционально) возможность работы не во всех 16, а в одном, двух, четырех и восьми подканалах — некий прообраз схемы множественного доступа OFDMA. Для этого каждый подканал и каждая группа подканалов имеет свой индекс (от 0 до 31).

Длительность полезной части T_b OFDM-символа зависит от ширины полосы канала BW и системной тактовой частоты (частоты дискретизации) F_s . $F_s \cdot T_b = N_{FFT}$. Соотношение $F_s/BW = n$ нормируется и в зависимости от ширины полосы канала принимает значения 86/75 (BW кратно 1,5 МГц), 144/125 (BW кратно 1,25 МГц), 316/275 (BW кратно 2,75 МГц), 57/50 (BW кратно 2 МГц) и 8/7 (BW кратно 1,75 МГц и во всех остальных случаях).

Канальное кодирование

Кодирование данных на физическом уровне включает три стадии — рандомизацию, помехоустойчивое кодирование и перемежение. Рандомизация — это умножение блока данных на псевдослучайную последовательность (ПСП), которую формирует генератор ПСП с задающим полиномом вида $1 + x^{14} + x^{15}$ (см. рис. 8.3).

В нисходящем потоке генератор ПСП инициализируется с началом кадра посредством кодового слова 4A80₁₆. Начиная со второго пакета кадра генератор ПСП инициализируется на основе идентификационного номера базовой станции BSID, идентификатора профиля пакета и номера кадра (рис. 8.10). В восходящем потоке все происходит аналогично, с той лишь разницей, что инициализация генератора ПСП по схеме, приведенной на рис. 8.10, происходит с первого пакета (вместо DIUC используется UIUC).

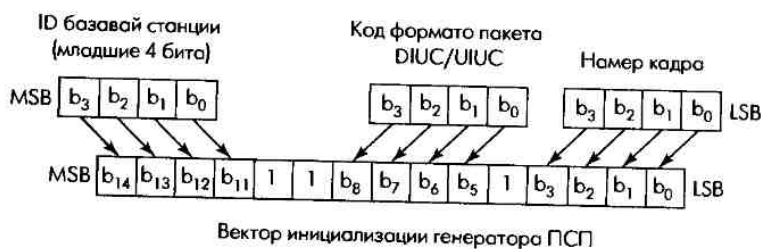


Рис. 8.10. Формирование вектора инициализации генератора ПСП для рандомизации нисходящего потока OFDM

Кодирование данных предполагает кодирование каскадным кодом, описанным в третьей главе, — внешним кодом Рида-Соломона и внутренним сверточным кодом. Алгоритм кодирования Рида-Соломона, используемый в данном стандарте, строится над полем Галуа GF (256). В базовом виде он оперирует блоками исходных данных по 239 байт, формируя из них закодированный блок размером 255 байт (добавляя 16 проверочных байт). Такой код способен исправить до 8 поврежденных байт или обнаружить до 15 поврежденных или стертых внутренним кодом байт. Поскольку реально используются блоки данных меньшей длины K , перед ними добавляются $(239-K)$ нулевых байт (так называемое укорочение линейного кода, см. главу 4). После кодирования эти байты удаляются. Если необходимо сократить число проверочных слов, так чтобы уменьшить число

восстанавливаемых байтов T , используются только $2T$ первых проверочных байт. Обязательные для поддержки в IEEE 802.16 варианты каскадного кодирования приведены в табл. 8.4.

Таблица 8.4. Обязательные схемы кодирования/модуляции в режиме OFDM

Модуляция	Блок данных до кодирования, байт	Кодер Рид-Соломона	Скорость кодирования сверточного кодера	Суммарная скорость кодирования	Блок данных после кодирования, байт
BPSK	12	(12,12,0)	1/2	1/2	24
QPSK	24	(32,24,4)	2/3	1/2	48
QPSK	36	(40,36,2)	5/6	3/4	48
16-QAM	48	(64,48,8)	2/3	1/2	96
16-QAM	72	(80,72,4)	5/6	3/4	96
64-QAM	96	(108,96,6)	3/4	2/3	144
64-QAM	108	(120,108,6)	5/6	3/4	144

После кодера Рид-Соломона данные поступают в сверточный кодер (рис. 8.11) с иницилирующими последовательностями (генераторами кода) $G_1 = 171_8$ (для выхода X) и $G_2 = 133_8$ (для Y) — так называемый стандартный код NASA. Его базовая скорость кодирования — $1/2$, т. е. из каждого входного бита он формирует пару кодированных бит X и Y . Кодовое ограничение этого кода $K = 7$. Упуская («выкалывая» или перфорируя) из последовательности пар элементы X_i или Y_i , можно получать различные скорости кодирования. Так, скорости $2/3$ соответствует последовательность $(X_1 Y_1 Y_2)$, скорости $3/4$ — $(X_1 Y_1 Y_2 X_3)$, $5/6$ — $(X_1 Y_1 Y_2 X_3 Y_4 X_5)$.

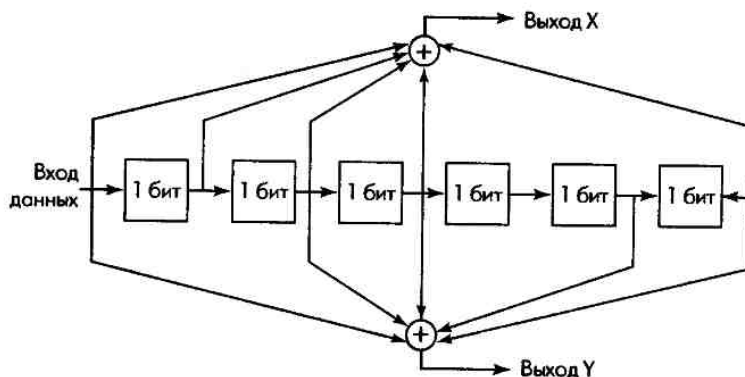


Рис. 8.11. Схема сверточного кодера

Кодер Рида-Соломона не используется с двухпозиционной модуляцией BPSK (например, при начальной инициализации АС или запросе полосы). Он также пропускается, когда используется часть субканалов OFDM. В этом случае скорость сверточного кодирования принимается равной общей скорости кодирования (см. табл.) (соответственно, размер исходного блока данных умножается на число используемых субканалов, деленное на 16).

Помимо описанного механизма кодирования, стандарт предусматривает опциональное применение блоковых турбокодов (основанных на кодах Хемминга и контроле четности) и сверточных турбокодов.

После кодирования следует процедура перемежения — перемешивания битов в пределах блока кодированных данных, соответствующего OFDM-символу. Эта операция проводится в две стадии. Цель первой — сделать так, чтобы смежные биты оказались разнесенными по несмежным несущим. На второй стадии смежные биты оказываются разнесенными в разные половины последовательности. Все это делается для того, чтобы при групповых (пакетируемых) ошибках в символе повреждались не смежные биты, которые легко восстановить при декодировании. Перемежение реализуется в соответствии с формулами

$$\begin{aligned} m_k &= (N_{cbps}/12) \cdot k \bmod 12 + \text{floor}(k/12); \\ j_k &= s \cdot \text{floor}(m_k/s) + (m_k + N_{cbps} - \text{floor}(12m_k/N_{cbps})) \bmod s; \\ k &= 0 \dots N_{cbps} - 1: \end{aligned}$$

где m_k и j_k — номер исходного k -го бита после первой и второй стадии перемежения соответственно; N_{cbps} — число кодированных бит в OFDM-символе (при заданном числе субканалов); s — 1/2 числа бит на несущую (1/2/4/6 бит для BPSK/QPSK/16-QAM/64-QAM соответственно, для BPSK $s = 1$). Функция $\text{floor}(x)$ — это наибольшее целое число, не превосходящее x ; функция $(x \bmod r)$ — остаток от x/r .

После перемежения начинается стадия модуляции. Исходя из выбранной схемы модуляции (BPSK/QPSK/16-QAM/64-QAM), блок представляется в виде последовательности групп бит, соответствующих модуляционным символам (по 1/2/4/6 бит). Каждой группе ставятся в соответствие значения Q и I из векторных диаграмм Грея (рис. 8.12), которые затем используются при непосредственной модуляции несущей. Для усреднения амплитуд квадратурных символов используются нормализованные значения Q и I , т.е. умноженные на коэффициенты c (для QPSK $c = \frac{1}{\sqrt{2}}$, для 16-QAM $c = \frac{1}{\sqrt{10}}$, для 64-QAM $c = \frac{1}{\sqrt{42}}$).

Пилотные несущие модулируются посредством BPSK. Значения сигналов на этих несущих определяются на основании ПСП w_k с задающим полиномом $x^{11} + x^9 + 1$, причем в нисходящем субкадре k — номер символа относительно начала кадра, в восходящем — номер символа относительно начала пакета (рис. 8.13). Инициализирующие слова генератора ПСП для нисходящего и восходящего потоков различны ($8FF_{16}$ и 555_{16} соответственно). Собственно значения BPSK-символов вычисляются как $c_{-88} = c_{-38} = c_{63} = c_{88} = 1 - 2w_k$; $c_{-63} = c_{-13} = c_{13} = c_{38} = 1 - 2\bar{w}_k$ в нисходящем канале и $c_{-88} = c_{-38} = c_{13} = c_{38} = c_{63} = c_{88} = 1 - 2w_k$; $c_{-63} = c_{-13} = 1 - 2\bar{w}_k$ — в восходящем. В результате получается так называемая ступенчатая конструкция OFDM, подробно описанная в конце главы 5.

Рис. 8.12. Векторные диаграммы Грея (представление модуляционных символов) для QPSK, 16-QAM и 64-QAM

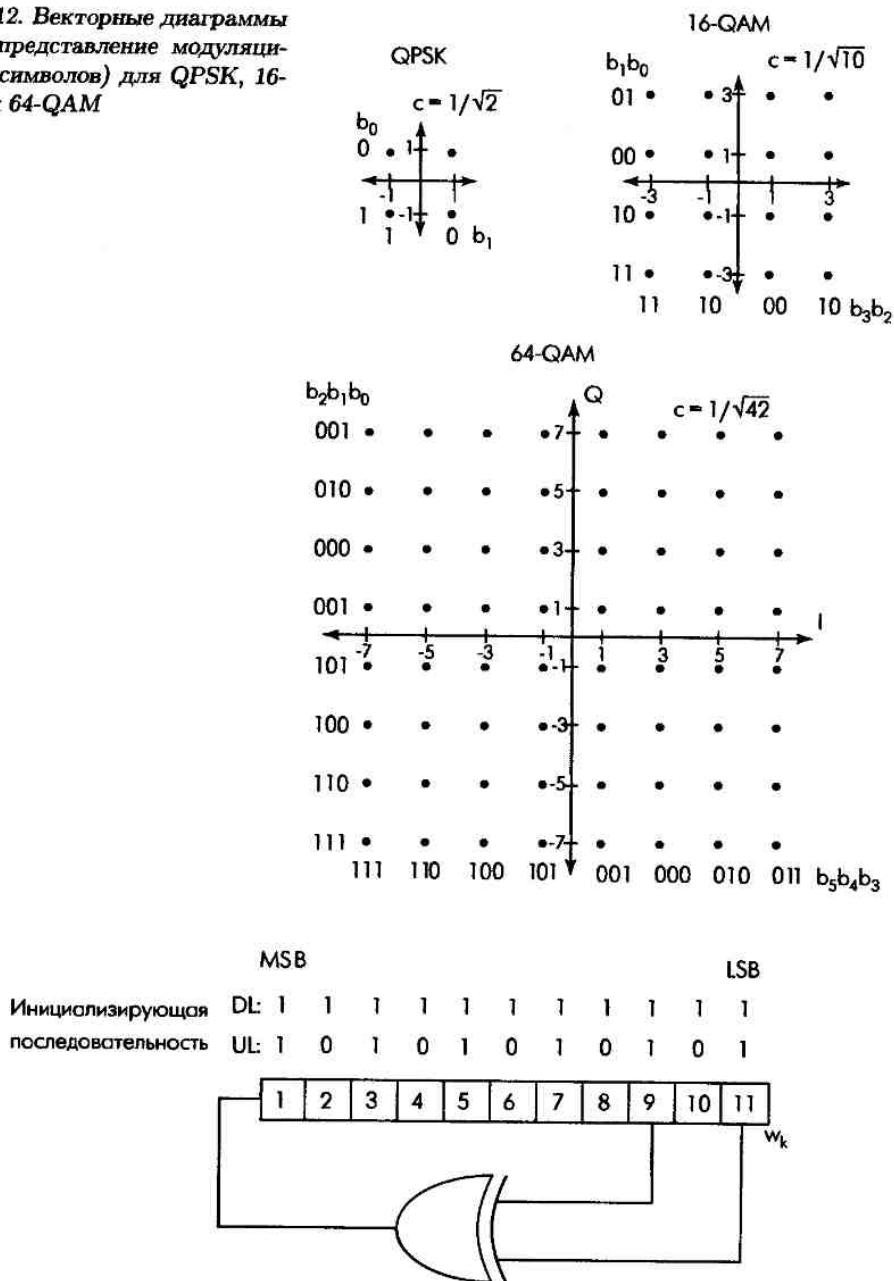


Рис. 8.13. Генерация модулирующей последовательности для пилотных несущих

После определения модуляционных символов посредством ОБПФ вычисляется сам радиосигнал и передается в передатчик. При приеме все процедуры происходят в обратном порядке.

Структура кадров

В режиме OFDM на физическом уровне для сетей с архитектурой «точка-многоточка» кадровая структура передачи принципиально мало чем отличается от режима SC. Так же как и в высокочастотной области, информационный обмен происходит посредством последовательности кадров (фреймов). Каждый фрейм (рис. 8.14) делится на два субкадра — нисходящий (*DL* — от БС к АС) и восходящий (*UL* — от АС к БС). Разделение на восходящий и нисходящий каналы — как временное (TDD), так и частотное (FDD). В последнем случае *DL* и *UL* транслируются одновременно, в разных частотных диапазонах.

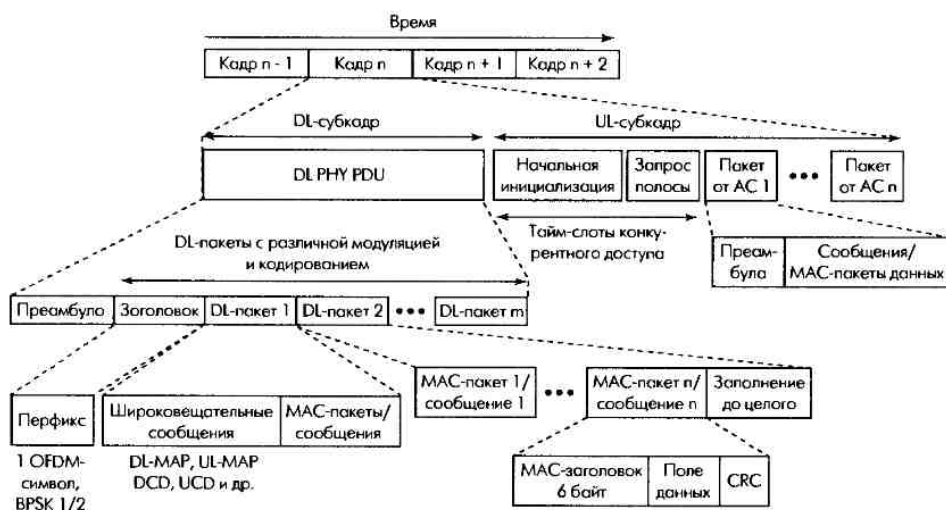


Рис. 8.14. Структура OFDM-кадров при временном дуплексировании

Нисходящий субкадр включает преамбулу, управляющий заголовок кадра (FCH — frame control header) и последовательность пакетов данных. Преамбула в нисходящем канале — посылка из двух OFDM-символов (длинная преамбула), предназначенная для синхронизации. Первый OFDM-символ использует несущие с индексами, кратными 4, второй — только четные несущие (модуляция — QPSK).

За преамбулой следует управляющий заголовок кадра — один OFDM-символ с модуляцией BPSK и стандартной схемой кодирования (скорость кодирования — $1/2$). Он содержит так называемый префикс кадра нисходящего канала (DLFP — Downlink Frame Prefix), который описывает профиль и длину первого (или нескольких начальных) пакета в DL-субкадре.

В первый пакет входят широковещательные сообщения (предназначенные всем АС) — карты расположения пакетов DL-MAP, UL-MAP, дескрипторы нисходящего/восходящего каналов DCD/UCD, другая служебная информация. Каждый пакет обладает своим профилем (схема кодирования, модуляция и т.д.) и передается посредством целого числа OFDM-символов. Точки начала и профили всех пакетов, помимо первого, содержатся в DL-MAP.

Нисходящий субкадр содержит интервал конкурентного доступа, включающий периоды для начальной инициализации АС (вхождение в сеть) и для запро-

са полосы передачи. Далее следуют временные интервалы, назначенные базовой станцией определенным абонентским станциям для передачи. Распределение этих интервалов (точки начала) содержится в сообщении UL-MAP. АС в своем временном интервале начинает трансляцию с передачи короткой преамбулы (один OFDM-символ, использует только четные несущие). За ним следует собственно информационный пакет, сформированный на MAC-уровне.

Длительность OFDM-кадров может составлять 2,5; 4; 5; 8; 10; 12,5 и 20 мс. Заданный базовой станцией, период построения кадров не может изменяться, поскольку в этом случае потребуются ресинхронизация всех АС.

Особенности запроса канальных ресурсов

Запрос на установление соединения не отличается от общепринятого в стандарте IEEE 802.16, за исключением дополнительного режима «концентрированного» запроса (Region-Focused). Он предназначен только для станций, способных работать с отдельными субканалами. В этом режиме в интервалах конкурентного доступа (заданных в UL-MAP) АС может передать короткий 4-разрядный код на одном из 48 субканалов, каждый из которых включает четыре несущие. Всего предусмотрено восемь кодов. Таблица кодов и подканалов приведена в тексте стандарта IEEE 802.16. Код и номера канала АС выбирает случайным образом.

Получив кодовое сообщение, БС предоставляет АС интервал для передачи «обычного» запроса на предоставление доступа (заголовок запроса MAC-уровня), если это возможно. Однако в отличие от других механизмов БС в UL-MAP не указывает идентификатор запросившей ее станции, а приводит номера кода запроса, подканала, а также порядковый номер интервала доступа, в течение которого было передан запрос. По этим параметрам АС и определяет, что интервал для запроса полосы передачи предназначен ей. Выбор момента для передачи 4-разрядного кода запроса доступа происходит случайным образом, по описанному выше алгоритму обращения к каналу конкурентного доступа.

Отметим, что в режиме OFDM канальный ресурс может предоставляться не только во временной области, но и в отдельных подканалах (группах подканалов), если БС и абонентские станции поддерживают такую возможность. Одно из наиболее важных применений такой опции — Mesh-сеть.

8.2.3 Mesh-сеть

Формально Mesh-сеть — это вид топологии сети IEEE 802.16 в режиме OFDM, и ее физический уровень — это OFDM. Однако различия Mesh-сети с уже рассмотренными режимами проявляются не только да и не столько на физическом уровне. Основное отличие Mesh-сети от рассматриваемой до сих пор архитектуры «точка-многоточка» в том, что если в последнем случае АС может общаться только с БС, то в Mesh-сети возможно взаимодействие непосредственно между АС. Поскольку сети стандарта IEEE 802.16 ориентированы на работу с широкими частотными каналами, Mesh-сети вошли в стандарт вовсе не в целях создания одноранговых локальных сетей — для этого есть стандарты группы IEEE 802.11. Причина в ином — необходим инструмент построения широкополосной сети, в которой трафик может передаваться по цепочке из нескольких станций,

ликвидируя тем самым проблемы передачи при отсутствии прямой видимости. Соответственно, и все механизмы управления, в принципе позволяющие построить децентрализованную распределенную сеть, ориентированы все же на древовидную архитектуру (корневой узел — БС) с выделенной базовой станцией и доминирующими потоками БС–АС.

Топология сети

В Mesh-сети все станции (узлы) формально равноправны. Однако практически всегда присутствует один узел, через который происходит обмен трафика Mesh-сети с внешним окружением (рис. 8.15). Такой узел называют базовой станцией Mesh-сети, именно на него возлагается часть необходимых для управления Mesh-сетью функций. При этом управление доступом может происходить либо на основе механизма распределенного управления, либо централизованным способом под управлением БС. Возможна и комбинация этих методов.

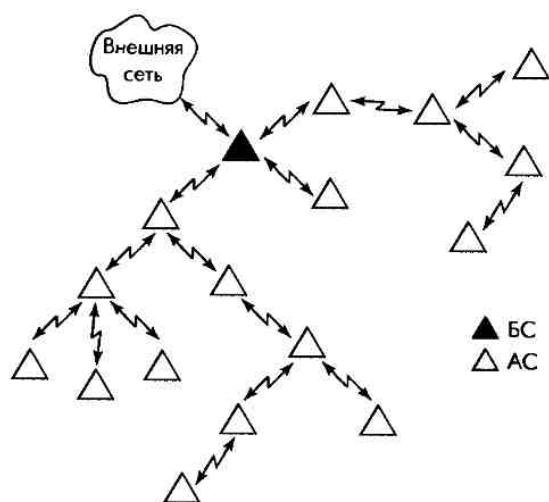


Рис. 8.15. Пример Mesh-сети

Базовое понятие в Mesh-сети — соседи. Под соседями определенного узла понимают все узлы, которые могут устанавливать с ним непосредственное соединение. Все они образуют соседское окружение. Узлы, связанные с заданным узлом через соседские узлы, называют соседями второго порядка. Могут быть соседи третьего порядка и т.д.

В Mesh-сети нет понятия восходящих/нисходящих каналов. Весь обмен происходит посредством кадров. Станции либо передают сообщения в отведенные им временные интервалы (в соответствии с предшествующим назначением каналов), либо получают доступ к каналам произвольным (конку-

рентным) образом. Каждый узел имеет уникальный 48-разрядный MAC-адрес. Кроме того, для идентификации внутри Mesh-сети станциям присваивается 16-разрядный сетевой идентификатор. Каждый узел постоянно хранит список данных обо всех своих соседях (с указанием удаленности, сектора для направленной антенны, примерной необходимой мощности передатчика для связи, задержки распространения сигнала и т.п.) и транслирует его в сеть с заданной периодичностью. На основании совокупности этих списков от каждого из узлов и происходит управление сетью.

Структура кадров и конфигурирование сети

Кадр Mesh-сети делится на управляющий субкадр и субкадр данных (рис. 8.16). Длина управляющего субкадра — переменная величина, задаваемая БС. Управ-

ляющий субкадр представляет собой набор пакетов MAC-уровня, с тем отличием, что сразу после общего заголовка MAC-пакета следует подзаголовок Mesh-сети. Управляющий субкадр, в зависимости от реализуемой функций, может быть двух типов — управления сетью (network control) и управления очередностью доступа к каналам связи (schedule control). В субкадрах всегда используется модуляция QPSK со скоростью кодирования 1/2.

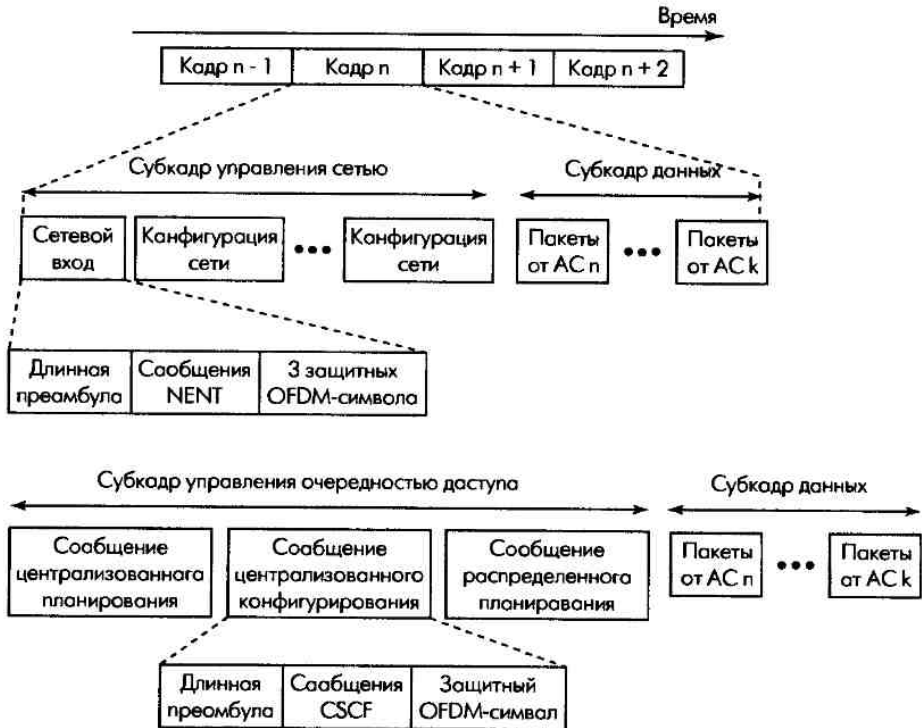


Рис. 8.16. Структура кадра Mesh-сети

Субкадры управления включают интервалы для подключения к сети новых устройств (Network entry — «сетевой вход») и следующие за ними сообщения «конфигурация сети». Сообщения типа «конфигурация сети» содержат всю необходимую информацию о составе сети. Они же реализуют процедуры управления. Эти сообщения генерирует каждый узел и транслирует по сети через свое соседское окружение. Среди передаваемой информации — списки соседей каждого узла, идентификационный номер БС и число ее соседей, номер логического канала для передачи графика доступа к каналам, удаленность узла (ранг соседства) от БС и т. д. Посредством таких сообщений с заданной периодичностью транслируется дескриптор сети — таблица, полностью описывающая текущие параметры сети. Она содержит такие параметры, как длительность кадров, длина управляющего субкадра, число интервалов для сообщений децентрализованного распределения ресурсов, периодичность следования субпакетов распределения ресурсов, профили пакетов, тип кодирования, соответствие логических каналов

физическим и т. п. Дескриптор сети передается от базовой станции ее соседскому окружению, от него — узлам со следующим рангом соседства и т. д. Периодичность передачи дескриптора сети нормирована.

«Сетевой вход» — это интервал, в течение которого новый узел может послать сообщение (NENT) о своем намерении подключиться к сети (аналог интервала конкурентного доступа в сети «точка-моготочка»). Перед этим он должен принять сообщение о конфигурации сети, выбрать узел для подключения, синхронизироваться с ним и лишь затем отправлять запрос. В ответ узел либо откажет в доступе, либо назначит новому узлу сетевой идентификатор, канал и временной интервал для проведения процедур аутентификации.

Методы управления каналными ресурсами

Распределение каналных ресурсов в Mesh-сети может быть централизованным и децентрализованным (распределенным). В свою очередь, децентрализованное распределение бывает координированным с БС и некоординированным.

Децентрализованное распределение ресурсов подразумевает, что распределение происходит в пределах одной группы соседей (т. е. между станциями, способными непосредственно связываться друг с другом). При координированном децентрализованном распределении узлы обмениваются между собой специальными сообщениями управления распределением (distributed scheduling — DSCH). Координированность заключается в том, что период выдачи таких сообщений каждой станцией определен и известен ее соседям. Координированные DSCH-сообщения передаются в субкадрах управления графиком доступа в оговоренных в сетевом дескрипторе интервалах. Некоординированные DSCH-сообщения передаются в субкадре данных.

DSCH-сообщения — это запросы на получение ресурса и ответные сообщения с предоставлением (подтверждением) свободного ресурса (временного интервала в субкадре данных). Ресурс предоставляется соседом под конкретное соединение.

Централизованное распределение ресурсов подразумевает древовидную топологию сети с БС в вершине. Оно реализовано посредством двух типов сообщений — централизованного конфигурирования CSCF и централизованного планирования CSCH. Эти управляющие сообщения размещаются в начале субкадра управления графиком доступа. Используя сообщения централизованного планирования CSCH, каждый узел определяет потребность в трафике своих дочерних узлов (т. е. трафик которых от/к БС проходит через данный узел) и сообщает свою потребность вышестоящему узлу — вплоть до БС. Проанализировав потребность, БС рассылает сообщение CSCH, информируя каждый узел о выделенной ему полосе пропускания (в бит/с) в восходящем и нисходящем направлениях. Исходя из этих данных каждый узел уже сам запрашивает (или назначает) расположение пакетов в субкадре данных у (для) своих соседских узлов посредством сообщений децентрализованного планирования DSCH.

Сообщения централизованного конфигурирования CSCF формируются БС и транслируются по сети для информирования всех ее узлов о текущем состоянии. CSCF включает такую информацию, как число доступных логических каналов и их перечень, перечень узлов в сети с указанием числа дочерних узлов для

каждого из них, а также профили восходящих/нисходящих пакетов для каждого дочернего узла.

8.2.4 Режим OFDMA

Режим OFDMA с точки зрения формирования модуляционных символов аналогичен OFDM. Различие проявляется в принципе разделения каналов. Один логический OFDMA-канал образован фиксированным набором несущих, как правило, распределенных по всему доступному диапазону частот физического канала. В упрощенном виде этот механизм опционально используется в режиме OFDM — вспомним разбиение канала на 16 подканалов.

Особенности формирования символов и канального кодирования

В OFDMA несущих значительно больше — 2048, соответственно, и число подканалов становится достаточным для организации работы сети: в разных режимах их от 32 до 70, по 24 или 48 информационных несущих в каждом. Системная тактовая частота — $8/7$ BW. Ширина физического канала BW не нормирована (в стандарте говорится «не менее 1 МГц»), но в реальных применениях вряд ли окажутся эффективными каналы менее 5 МГц.

Метод формирования, структура OFDM-символов и механизм канального кодирования в OFDMA схожи с описанным для OFDM. Канальное кодирование включает рандомизацию, помехоустойчивое кодирование, перемежение и модуляцию. Метод рандомизации также практически идентичен OFDM (различны способы формирования инициализирующего вектора генератора ПСП).

Помехоустойчивое кодирование в OFDMA в качестве обязательного предусматривает только сверточный кодер — такой же, как в OFDM, и с тем же набором скоростей кодирования. Кодера Рида-Соломона нет. Опционально предусмотрено применение блочных и сверточных турбокодов. Метод перемежения также идентичен с OFDM — надо лишь в соответствующих формулах заменить 12 (для OFDM) на 16.

Схемы модуляции несущих полностью совпадают с OFDM, с той лишь разницей, что предусмотренный набор включает только QPSK и 16-QAM со скоростями кодирования $1/2$ и $3/4$, а также опционально 64-QAM со скоростями кодирования $1/2$, $2/3$ и $3/4$. Однако в OFDMA после формирования символов квадратурной модуляции и усреднения их амплитуд (нормировке на параметр c) последовательность символов на каждой несущей умножается на бинарную ПСП w_k , схема генератора которого идентична приведенной на рис. 8.13. Каждая k -я несущая умножается на значение $1 - 2w_k$ (т. е. если $w_k = 0$, то $1 - 2w_k = 1$ и символ не изменяется; если $w_k = 1$, символ умножается на -1). Символы пилотных несущих модулируются методом BPSK, их значения вычисляются как $c_k = 1 - 2w_k$. Однако, поскольку мощность сигналов пилотных несущих в нисходящем канале (опционально — и в восходящем) должна быть на 2,5 дБ выше средней мощности информационных несущих, значение c_k дополнительно умножается на $4/3$.

Структура кадров, методы распределения несущих

Структура кадров (рис. 8.17) в OFDMA схожа со всеми рассмотренными режимами в том, что сохраняется подразделение на восходящий и нисходящий субкадры, как временное, так и частотное. Длительность кадра может составлять 2; 2,5; 4; 5; 8; 10; 12,5 и 20 мс. Кадр — это последовательность OFDMA-символов. Каждый OFDMA-символ включает набор субканалов. Пакеты данных могут передаваться одновременно, на различных OFDMA-подканалах.

Для описания структуры кадра в OFDMA используется понятие слота — минимального ресурса для передачи данных. Слот занимает один подканал и от одного до трех последовательных OFDMA-символов. В нисходящем субкадре длительность слота — один символ в режиме FUSC, два — в режиме PUSC; в восходящем субкадре длительность слота всегда равна трем OFDMA-символам.

Подканал — это набор несущих частот (как и в OFDM). Распределение несущих по подканалам, равно как и число несущих на один подканал, зависит от направления передачи и метода распределения несущих. Стандарт IEEE 802.16 описывает несколько способов распределения несущих, как в нисходящем канале, так и в восходящем. Принципиально они подразделяются на FUSC (full usage of the subchannels) — полное использование подканалов передатчиком БС и PUSC (partial usage of subchannels) — использование групп подканалов (сегментов), т. е. не всего доступного диапазона. Какие именно подканалы используются в режиме PUSC, однозначно определяют номера сегментов.

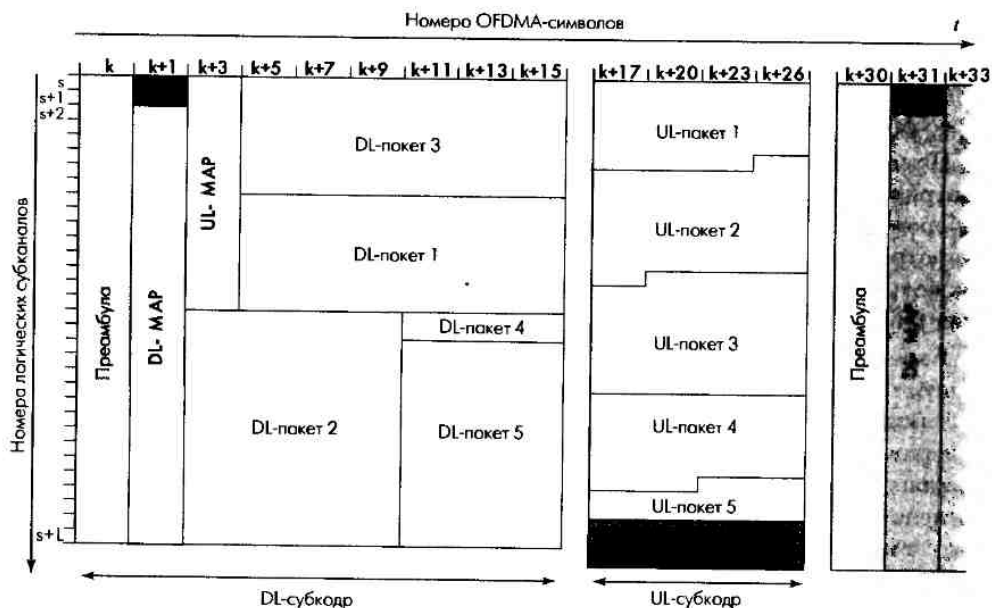


Рис. 8.17. Структура OFDMA-кадров

В методах PUSC и FUSC (и их вариациях) одному субканалу присваиваются несущие, равномерно распределенные по всему доступному физическому кана-

лу. Используется и другой подход — применение в подканалах набора последовательных соседних частот. Он реализован в методе AMC (Advanced modulation and coding), предназначенном для работы с адаптивными антенными системами.

В методе AMC используется также 2048 несущих. Из них 160 нижних и 159 верхних образуют защитные интервалы, центральная частота не используется. Оставшиеся несущие последовательно разбиваются на 192 группы (группа именуется bin) по 9 несущих в каждой. Центральная (пятая) частота в каждой группе — пилотная. Четыре соседние (по частоте) группы образуют полосу (band). Шесть групп в одной полосе образуют AMC-подканал. Понятие подканала в данном случае частотно-временное, т. е. подканал — это несколько (например, два) OFDMA-символов и три несущие. Структура AMC-подканалов в восходящем и нисходящем субканалах одинакова.

Отметим, что в пределах одного субканала возможно использование различных механизмов распределения несущих по подканалам — FUSC, PUSC, AMC и т.д. Границы соответствующих зон (именуемых в стандарте зонами перестановки — permutation zone) определены в картах субканалов (рис. 8.18).



Рис. 8.18. Совмещение различных «зон перестановки» в OFDMA-кадре

Нисходящий OFDMA-канал

В нисходящем канале (рис. 8.17) первый символ — это преамбула. Несущие преамбулы модулируются посредством BPSK специальным псевдослучайным кодом, зависящим от используемого сегмента (в режиме PUSC) и переменной IDcell, задаваемой на MAC-уровне. Одному сегменту с номером $n = [0 \dots 2]$ соответствует набор несущих с шагом 3 и начальным сдвигом n . Распознав тип преамбулы, АС сразу определяет значение переменной IDcell и режим работы БС.

За преамбулой следуют два символа, передающие заголовок кадра FCH и карту распределения нисходящего поля нисходящего канала DL-MAP. Заголовок транслируется посредством QPSK со скоростью кодирования 1/2. Он содержит префикс нисходящего канала (DL Frame prefix), в котором указываются используемые сегменты, и параметры карты нисходящего канала (длина, используемый

метод кодирования и число повторений), транслируемой сразу за заголовком кадра. Также в заголовке используется флаг, установка которого означает изменение в расположении области конкурентного доступа в восходящем субкадре по отношению к предыдущему кадру.

Далее транслируется карта восходящего канала UL-MAP и нисходящие пакеты данных для разных АС.

Режим FUSC означает, что используется весь диапазон физического канала (все возможные несущие). Это 1702 несущие частоты, еще 173 и 172 несущих вверх и вниз диапазона, соответственно, образуют защитные интервалы. Центральная частота с индексом 1024 не используется.

В режиме FUSC прежде всего назначаются пилотные частоты. Они подразделяются на фиксированные и переменные. Списки тех и других приведены в стандарте. Термин «переменные пилотные частоты» означает, что в каждом четном OFDMA-символе их индексы соответствуют приведенным в документе IEEE 802.16, в каждом нечетном — увеличиваются на 6 (нумерация OFDMA-символов начинается с 0). Всего предусмотрено 166 пилотных частот, из них 24 — фиксированные. И фиксированные, и переменные пилотные частоты разбиты на два набора, одинаковые по объему. Это разбиение имеет значение только при работе с адаптивными антенными системами в режиме пространственно-временного кодирования (STC).

После назначения пилотных частот оставшиеся 1536 несущих предназначены для передачи данных. Они подразделяются на $N_{subchannels} = 32$ подканала по $N_{subcarriers} = 48$ несущих в каждом. Назначение информационных несущих подканалам происходит в соответствии с формулой

$$subcarrier(k, s) = N_{subchannels}n_k + \{P[(s + n_k) \bmod N_{subchannels}] + ID_{cell}\} \bmod N_{subchannels}, \quad (8.1)$$

где $subcarrier(k, s)$ — индекс несущей k в подканале s , $s = [0 \dots N_{subchannels} - 1]$, $k = [0 \dots N_{subcarriers} - 1]$, $n_k = (k + 13s) \bmod N_{subcarriers}$. ID_{cell} — определяемый на MAC-уровне идентификатор отдельного сегмента БС (задаваемая базовой станцией целая переменная в диапазоне 0–31). $P(x)$ означает x -й элемент последовательности перестановок $\{P\}$, приведенной в стандарте ($P = \{3, 18, 2, 8, 16, 10, 11, 15, 26, 22, 6, 9, 27, 20, 25, 1, 29, 7, 21, 5, 28, 31, 23, 17, 4, 24, 0, 13, 12, 19, 14, 30\}$). Операция $x \bmod k$ означает остаток от x/k .

Очевидно, что перед применением приведенной формулы информационные несущие должны быть перенумерованы так, чтобы их индексы укладывались в диапазон 0–1535 (последнее значение соответствует физическому индексу 1702) F, т. е. пронумерованы подряд, без учета пилотных частот. Поскольку в четных и нечетных символах расположение пилотных частот различно, распределение информационных несущих для них также нужно вычислять независимо.

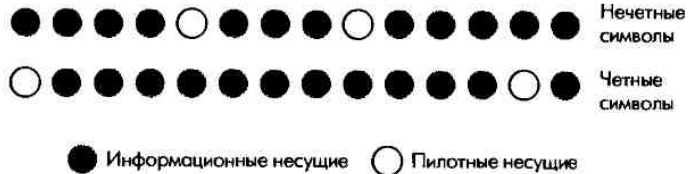
В режиме PUSC весь доступный диапазон подканалов делится на ряд независимых сегментов. Для работы используется часть из них, но не менее 12 подканалов. Всего в режиме PUSC может быть 60 подканалов. Они разбиты на шесть неравных сегментов. Из них три базовых (сегменты 0, 1 и 2), каждый включает 12 каналов (0–11, 20–31 и 40–51 подканалы соответственно).

Символ в режиме PUSC формируется по следующему принципу. Всего предусмотрено 2048 частот, из них центральная (с индексом 1024) и защитные (184 нижние и 183 верхние) не используются. Оставшиеся 1680 несущих последовательно разбивают на 120 кластеров, каждый содержит 14 несущих. После чего последовательные физические кластеры перенумеровываются в «логические» в соответствии с формулой

$$\text{LogicalCluster} = \text{RenumberingSequence}((\text{PhysicalCluster} + 13\text{IDcell}) \bmod 120),$$

где $\text{RenumberingSequence}(x)$ — соответствующий элемент приведенной в стандарте IEEE 802.16 последовательности перестановок, IDcell — определяемый на MAC-уровне идентификатор отдельного сегмента БС (задаваемая базовой станцией целая переменная в диапазоне 0–31). Эта операция фактически означает перемешивание — распределение последовательных групп несущих по всему диапазону физического канала. Далее логические кластеры разбиваются на шесть групп (0–23, 24–39, 40–63, 64–79, 80–103, 104–119) по 24 и 16 кластеров. Большие группы соответствуют большим сегментам (по умолчанию, группа 0 соответствует сегменту 0, группа 2 — сегменту 1, группа 4 — сегменту 2). В каждом кластере определяются пилотные несущие: для четных символов это 5-я и 9-я несущие, для нечетных — 1-я и 13-я (рис. 8.19).

Рис. 8.19. Структура кластера



Таким образом, набору подканалов в пределах сегмента или нескольких сегментов оказывается поставленным в соответствие набор несущих (для 12 подканалов — 336 несущих, из них 24 пилотные и 288 информационных). Информационные несущие в сегменте нумеруются подряд, не учитывая пилотные частоты, после чего в соответствии с формулой (8.1) каждому подканалу назначаются по 24 несущих. В данном случае в формуле (8.1) используются значения $N_{\text{subchannels}} = 12$ или 8, $N_{\text{subcarriers}} = 24$, а также специальные перестановочные последовательности P_{12} и P_8 для сегментов из 12 и 8 каналов соответственно (приведены в стандарте).

Кроме рассмотренных методов распределения несущих, в стандарте предусмотрены и опциональные механизмы, в частности так называемый optional FUSC, принципиально не отличающийся от рассмотренного.

Восходящий канал

Восходящий субкадр следует непосредственно за нисходящим через интервал TTG. Он содержит пакеты от АС и интервал для запроса доступа/инициализации. Минимальный размер одного сообщения в восходящем подкадре (слот) — 3 OFDMA-символа в одном подканале. Это привело к появлению в документе IEEE 802.16 термина «фрагмент» (мозаичный элемент, tile).



Рис. 8.20. Структура «фрагмента» восходящего канала

Фрагмент представляет собой совокупность трех символов и четырех несущих, в котором положения пилотных частот жестко определены (рис. 8.20).

Весь частотный диапазон канала (1680 несущих) разбивается на 420 последовательных фрагментов, по четыре несущие в каждом. Предусмотрено 70 подканалов. Каждый из них включает шесть фрагментов, т. е. 24 несущие на символ в одном подканале. Распределение фрагментов по подканалам происходит следующим образом. Все 420 фрагментов разбиваются на шесть групп по 70 фрагментов. В каждый подканал включается по одному фрагменту из каждой группы в соответствии с уравнением

$$Tile(n, s) = 70n + \{P[(n + s) \bmod 70] + UL_IDcell\} \bmod 70,$$

где $Tile(n, s)$ — фрагмент n подканала s , $n = [0 \dots 5]$, $s = [0 \dots 69]$, $P(x)$ — перестановочная последовательность, UL_IDcell — переменная в диапазоне 0–69, задаваемая БС на MAC-уровне. В результате каждому подканалу в каждом символе назначается свой набор несущих.

После распределения по подканалам происходит нумерация информационных несущих в каждом слоте — всего их в трех символах 48. Информационные частоты нумеруются в подканале начиная с наименьшей несущей фрагмента с наименьшим индексом первого символа — сначала в первом символе, затем во втором и третьем. После чего информационные несущие в каждом слоте перенумеровываются в соответствии с формулой $subcarrier(n, s) = (n + 13s) \bmod 48$, где s — номер подканала, $n = [0 \dots 47]$ (т. е. происходит циклический сдвиг нумерации информационных несущих на $13s$ в каждом подканале s).



Рис. 8.21. Структура «фрагмента» восходящего канала в опциональном режиме

Отметим, что в тексте документа IEEE 802.16 происходит подмена терминов: подканалом в восходящем субкадре авторы текста IEEE 802.16 называют именно слот, информационную структуру размером 24 несущих на три символа. И когда в документе написано, что в субканале 48 информационных несущих, следует помнить, что с точки зрения правильной терминологии речь идет не о субканале, а о слоте. Реальных же несущих (т. е. физических частот) в субканале всего 24. Умножая их на 3 (число OFDMA-символов в слоте) и вычитая 24 пилотные несущие, как раз и получим 48 информационных несущих.

Опционально в восходящем канале предусмотрен режим, в котором во фрагменте одна пилотная частота (рис. 8.21), шесть фрагментов на подканал, всего 96 подканалов (1728 используемых частот).

Запрос полосы и регистрация в сети

Механизмы запроса начальной инициализации в сети и первичного запроса полосы пропускания в режиме OFDMA схожи и принципиально отличаются от

других режимов. Для этих запросов в OFDMA используется специально выделенный канал. Он назначается БС и состоит из шести последовательных подканалов, индексы которых приведены в UL-MAP. Запрос представляет собой 144-разрядный CDMA-код, передаваемый посредством BPSK, т. е. 1 бит на несущую. В результате для передачи такого кода достаточно шести подканалов (24 информационных несущие в каждом). Сам код формируется в генераторе ПСП — 15-разрядном сдвиговом регистре с задающим полиномом $1 + X^1 + X^4 + X^7 + X^{15}$. Старшие шесть разрядов вектора инициализации генератора ПСП равны переменной UL_IDcell , остальные девять — константа. Номер кода определяется начальной точкой (т. е. числом тактов генератора ПСП после инициализации) — всего предусмотрено 256 кодов. Причем БС использует только часть из всех возможных кодов — сначала N кодов начальной инициализации, за ними следуют M кодов периодического определения параметров АС, далее L кодов запроса полосы. Для каждой БС задается точка начала этой последовательности ($N + M + L$).

Начальная инициализация происходит так: АС, приняв дескриптор восходящего канала и UL-MAP, определяет набор CDMA-кодов и посылает в отведенном интервале случайно выбранный код из группы возможных. Один и тот же код транслируется в двух последовательных OFDMA-символах. Если продолжительность канала конкурентного доступа составляет более одного слота, АС может отправить CDMA-код в четырех последовательных символах, причем коды должны быть смежными (т. е. следовать в ПСП один за другим).

Успешно приняв и распознав CDMA-код (а это может и не произойти, поскольку в интервале конкурентного доступа возможны коллизии при одновременной работе передатчиков нескольких АС), базовая станция не знает, от какой АС пришел запрос. Поэтому в ответ в UL-MAP следующего кадра она указывает номер принятого CDMA-кода, субканал и символ, в котором код был отправлен. Так АС определяет, что именно ее запрос принят, и понимает, что следующее за этим широкополосное сообщение с указанием диапазона запроса (номера символа, подканала и длительности) предназначено именно ей. В этом сообщении БС передает необходимые параметры для процесса инициализации в сети (включая идентификатор соединения CID, присвоенный MAC-адрес, набор физических параметров и др.). Далее в указанный в UL-MAP интервал АС приступает к штатной процедуре регистрации в сети.

Первичный запрос полосы в методе OFDMA может происходить двумя способами — посредством заголовков запроса полосы, как и в остальных режимах, и путем посылки CDMA-кода запроса полосы в интервале конкурентного доступа. Посылка кода запроса полосы (равно как и кода периодического измерения параметров) происходит в одном OFDMA-символе. Возможна и посылка трех последовательных кодов в трех символах (какой из вариантов необходимо использовать, указывается в UL-MAP). Приняв CDMA-код, БС в UL-MAP повторяет его номер и параметры, а также сообщает интервал для отправки заголовка запроса полосы уже обычным способом.

8.3 Поддержка адаптивных антенных систем

Важнейшая особенность стандарта IEEE 802.16, принципиально отличающая его, скажем, от стандартов IEEE 802 a/b/g, — это наличие встроенных средств под-

держки адаптивных антенных систем (AAS). Подробно подобные системы описаны в главе 9. Разумеется, применение AAS — не обязательное требование стандарта. AAS — это системы с секторными направленными антеннами (метод формирования диаграмм направленности антенн в стандарте не оговаривается), т. е. антенные системы с несколькими антенными элементами. Применение AAS существенно увеличивает потенциальную емкость сети стандарта IEEE 802.16, поскольку в разных секторах БС возможна работа в одних и тех же каналах (частотных и OFDMA). Кроме того, направленные антенны позволяют существенно уменьшать общую излучаемую мощность. В результате снижается и межканальная интерференция. Не менее важно применение многоэлементных антенных систем для улучшения прохождения сигналов в каналах с замираниями, так называемых методов пространственно-временного кодирования (разнесения), STC.

Поддержка ASS в спецификации IEEE 802.16 означает модификацию протоколов на физическом и MAC-уровнях, наличие специальных управляющих и контролирующих сообщений для работы с адаптивными антеннами.

8.3.1 Работа с направленными AAS

Стандарт допускает в рамках одного кадра транслировать как ненаправленный, так и направленный (посредством AAS) трафик (рис. 8.22). Для разграничения зон не-AAS и AAS-трафика используются специальные сообщения. Принцип работы с AAS в режимах OFDM и OFDMA, равно как и в SCa, достаточно схож. Наиболее полно он описан в стандарте для случая OFDMA, поэтому остановимся именно на нем.

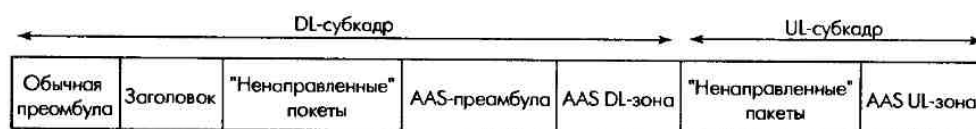


Рис. 8.22. Структура кадров с зоной AAS

Механизм Diversity-Map Scan

В режиме OFDMA предусмотрено два метода работы с AAS — с распределенными несущими в подканале (FUSC, PUSC) и с последовательными несущими (AMC). Каждый из методов в начале AAS-зоны предусматривает передачу OFDMA-символа преамбулы AAS-зоны и заголовка с префиксом AAS-зоны. Для передачи этих сообщений в AAS-зоне нисходящего субкадра выделены специальные подканалы (два старших для FUSC/PUSC и четвертый с начала и четвертый с конца подканалы в AMC). Сообщения в этих подканалах могут повторяться несколько раз, с тем что, если используется не широкополосная трансляция, а передача с переключением лучей, сообщения с префиксом допили бы до всех AC. В префиксе указывается код луча антенны, тип и размеры преамбулы ASS-зоны (в восходящем и нисходящем каналах), область для начальной инициализации/запросов полосы, а также области в кадре для каждого AAS-соединения. Префикс, как и в штатном режиме, передается посредством QPSK

со скоростью кодирования $1/2$ и двукратным повтором (в пределах одного символа). Основное назначение префикса — сообщить АС о том, как будут переданы карты UL/DL-каналов для разделенных по направлениям лучей групп пользователей (очевидно, что распределение канальных ресурсов может происходить независимо в каждом луче).

Для работы в режиме AMC-AAS кадры могут объединяться в суперкадр длительностью не менее 20 обычных кадров. В суперкадр входит по крайней мере один широкополосный кадр, содержащий дескрипторы и карты DL/UL-каналов. Смысл такого объединения — обеспечить минимум управляющих сообщений для группы кадров.

Метод Direct Signaling

Перечисленные методы работы с AAS используют так называемый механизм Diversity-Map Scan — сканирование (абонентскими станциями) разнесенных карт распределения канальных ресурсов. В режиме OFDMA предусмотрен и другой способ работы с AAS — метод прямой сигнализации (Direct Signaling Method). Он использует механизм последовательного распределения несущих AMC.

Особенность метода — в каждом кадре в AAS-зоне выделяется от одного до четырех каналов доступа/распределения ресурсов (BWAA — bandwidth allocation /access). Каждый BWAA-канал состоит из двух субканалов, расположенных в верхней и нижней частях диапазона симметрично относительно центральной частоты (если BWAA-канал один, то он включает самый верхний и самый нижний подканалы). В этом канале передаются префикс нисходящего субканала (для режима Direct Signaling Method), карты UL-MAP и DL-MAP для каждой из пространственно разделенных АС или групп АС. Благодаря точной пространственной настройке AAS данный метод позволяет в одном кадре передавать сообщения множеству пользователей.

В методе прямой сигнализации предусмотрены четыре специальных кодовых сообщения — обучения обратного соединения RLT (reverse link training), доступа в обратном соединении RLA (reverse link access), обучения прямого соединения FLT (forward link training) и инициирования прямого соединения FLI (forward link initiation). Первые два сообщения использует АС, вторые два — БС. Для начальной инициализации или запроса полосы АС посылает сообщение RLA в канале BWAA. Оно предшествует сообщениям запроса полосы или начального доступа и используется БС для точной настройки своей антенной системы на данную АС. В ответ БС передает сообщение FLI — уникальный код для каждой АС (БС может сама инициировать соединение, послав FLI). FLI транслируется в подканале, выделенном для данной АС. Каждая абонентская станция сканирует все подканалы и, обнаружив по кодовой последовательности адресованное ей сообщение начальной инициализации, отправляет в ответ в том же самом канале (в отведенном для нее временном интервале) последовательность RLT, предназначенную для точной настройки антенн БС на АС в данном подканале. В результате, выполнив все необходимые подстройки, БС и АС устанавливают соединение, в течение которого происходит обмен данными. Причем пакетам дан-

ных предпшествоуют тренировочные последовательности FLT (со стороны БС) и RLT (со стороны АС).

8.3.2 Пространственно-временное кодирование

Еще одна важная особенность применения многоэлементных антенных систем — это возможность использовать пространственно-временное разнесение передающих каналов (Space-Time Coding, STC) для улучшения прохождения радиосигналов. Идея метода — разнести, пространственно и во времени, источник одного и того же сигнала, т. е. несколько изменить условия его прохождения. Вероятность безошибочного приема такого сигнала (после соответствующей первичной обработки в приемнике) существенно возрастает.

В стандарте IEEE 802.16 используется схема пространственно-временного разнесения, предложенная Аламоути [118]. Суть метода проста — выходной поток символов разбивается на два (например, четные и нечетные символы), формируемые параллельно (рис. 8.23). В передатчике используется два антенных канала, действующих параллельно и использующих общий тактовый генератор (что обеспечивает синхронность). Таким образом, реализуется так называемая схема канала MISO (Multiple Input Single Output) — несколько входов и один выход (по отношению к каналу).

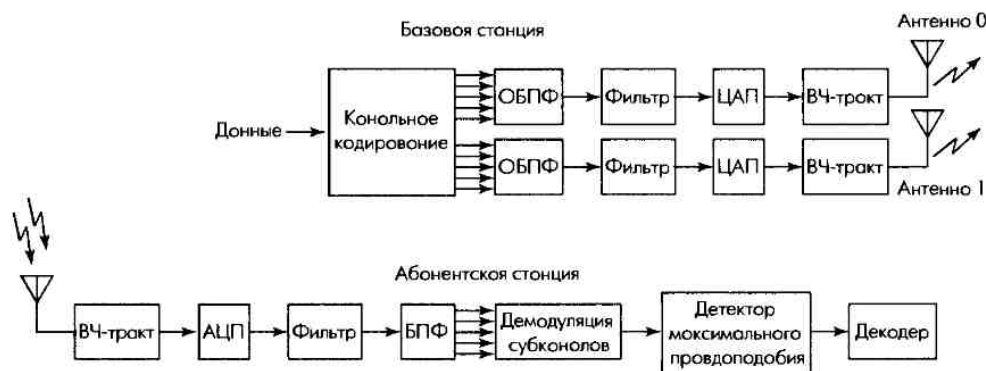


Рис. 8.23. Метод пространственно-временного кодирования по схеме MISO

Сначала антенна 0 транслирует символ S_0 , антенна 1 — символ S_1 . В следующий символьный интервал антенна 0 передает символ $-S_1^*$, антенна 1 — символ S_0^* (S^* означает комплексное дополнение к S). Приемник работает с одной антенной и в каждом символьном интервале принимает сигналы r_0 и r_1 . Зная передаточные характеристики каналов (h_0 и h_1), в приемнике можно восстановить переданные сигналы S_0 и S_1 согласно формулам (разумеется, вычисленные значения являются некоторым приближением к исходным значениям S_0 и S_1):

$$S_0 = h_0^* r_0 + h_1 r_1^*; \quad S_1 = h_1^* r_0 - h_0 r_1^*.$$

С точки зрения протоколов физического уровня применение STC не требует особых действий. Зона, транслируемая посредством STC, помечается в DL-MAP каждого кадра.

Особенности STC в методе OFDMA

В методе OFDMA предусмотрен ряд особенностей реализации STC. Прежде всего при формировании OFDMA-символов набор пилотных частот зависит от четности символа и номера антенного канала. Кроме того, поскольку в OFDMA разнесению используется и частотное, в дополнение к пространственно-временному разнесению используется и частотное, посредством частотных скачков (frequency hopping diversity coding — FHDC). Суть данного механизма: допустим, несущие в субканале X модулируются сигнальным вектором S_0 , в субканале $X + 1$ — вектором S_1 . Именно такой сигнал передает антенна 0. Антенна 1 транслирует сигнал, в котором несущие подканала X модулируются вектором $-S_1^*$, несущие подканала $X + 1$ — вектором S_0^* . Восстановление в приемнике происходит аналогично рассмотренному варианту STC, только вместо передаточных характеристик двух антенных каналов используются характеристики, связанные с субканалами X и $X + 1$. Под принятыми сигналами r_0 и r_1 понимают принятые сигналы в субканалах X и $X + 1$ соответственно. Из них восстанавливают S_0 и S_1 .

Очевидно, что данную методику можно перенести на пары субканалов, т.е. все подканалы OFDMA-символа разбиваются на смежные пары ($X, X + 1; Y, Y + 1; \dots$). В антенне 0 они передаются без изменений, в антенне 1 в каждой паре происходит описанное преобразование.

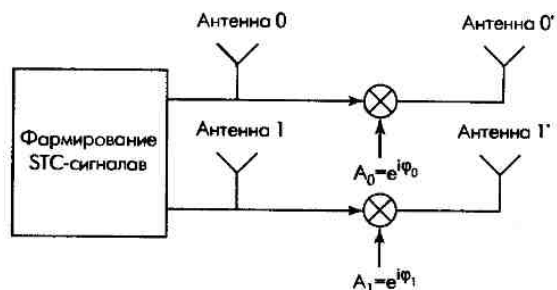
Все изложенные схемы преобразования можно описать матрицей

$$A = \begin{bmatrix} S_1 & -S_2^* \\ S_2 & S_1^* \end{bmatrix}.$$

Однако возможна и упрощенная схема: $B = \begin{bmatrix} S_1 \\ S_2 \end{bmatrix}$, обеспечивающая, однако, двукратный выигрыш в скорости. Вид матрицы преобразования задается базовой станцией в картах соответствующих каналов.

Метод OFDMA допускает применение STC/FHDC не только в нисходящем, но и в восходящем канале. Кроме того, возможно применение STC на базе не только двух, но и четырех антенных элементов. В последнем случае помимо базовых антенн 0 и 1 (рис. 8.24) добавляются антенны 0' и 1', сигнал в которых смещен по фазе (например, сигнал в антенне 0' $S'_0 = S_0 \cdot e^{i\varphi_0}$).

Рис. 8.24. Схема STC с четырьмя передающими антеннами



8.4 Аппаратная поддержка стандарта IEEE 802.16

Выше был кратко рассмотрен стандарт IEEE 802.16. Из изложения видно, что спецификация IEEE 802.16 предоставляет создателям аппаратуры достаточно широкие возможности, не оговаривая при этом конкретные способы реализации предусмотренных стандартом алгоритмов и механизмов. Фактически IEEE 802.16, как и положено стандарту, описывает самые общие правила игры, следуя которым возможно производить совместимую аппаратуру. Если эта цель будет достигнута, нас ожидает небывалое оживление рынка средств беспроводного доступа, а также связанных с ним услуг. Доказательство тому — имена лишь некоторых членов WiMax Forum, таких как Intel, Nokia, Analog Devices, Atheros Communications, Fujitsu Microelectronics America и многие другие.

Более того, в низкочастотной области (менее 11 ГГц) некоторые производители объявили о создании соответствующей элементной базы. Именно в этом диапазоне события могут разворачиваться более стремительно. Так, по данным исследования компании Yankee Group, число абонентов систем MMDS (эту систему можно рассматривать как составляющую IEEE 802.16) к 2006 году в США составит около 900 тыс. (в 45 раз больше, чем в 2000 году) [231].

8.4.1 Интегральная элементная база

Видимо, первой свой продукт на этот рынок выпустила канадская фирма Wavesat (www.wavesat.com), еще в феврале 2003 года анонсировавшая сверхбольшую интегральную схему (СБИС) OFDM-модема DM110 (рис. 8.25). Он был реализован в ASIC, т. е. методом заказного проектирования, в корпусе типа BGA-1156. Схема обладала возможностями, несколько отличающимися от требований IEEE 802.16 (в чем-то избыточными). Так, поддерживалась модуляция 4-, 16- и 64-QAM, 1024-точечное БПФ, временное и частотное дуплексирование, формирование OFDM-пакетов с защитным кодированием. Допустимая ширина канала — от 3 до 7 МГц, скорость передачи данных — до 35 Мбит/с. Напряжение питания ядра и периферии составляло 2,5 и 3,3 В соответственно.

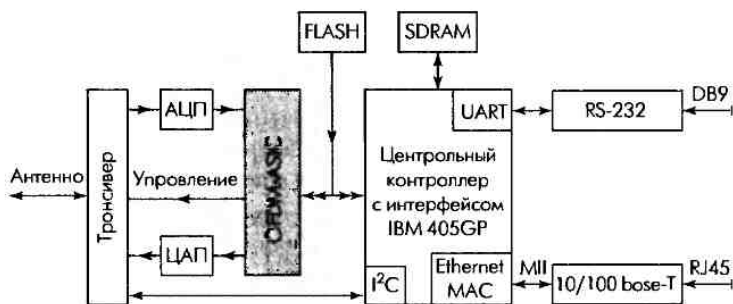


Рис. 8.25. Схема устройства на базе OFDM-модема DM110 компании Wavesat

В декабре 2004 года фирма Wavesat объявила о начале продаж своего нового OFDM-модема — СБИС DM256. СБИС реализована в корпусе PQFP-208 и принципиально отличается от предшественницы. DM256 оснащена ЦАП и АЦП

(10 разрядов). Входной/выходной интерфейс реализованы в виде как квадратурных составляющих (I и Q), так и модулированного сигнала на промежуточной частоте 10 МГц. Поддерживается модуляция 2/4/16/64-QAM. В микросхеме реализованы разработанные компанией механизмы временной и частотной синхронизации, поддерживается временное и частотное разделение каналов, в последнем случае — дуплексный и полудуплексный режимы. Ширина канала — 1,75; 3; 7 и 10 МГц, длительность защитного интервала — от 1/4 до 1/32 от длительности OFDM-символа. На аппаратном уровне поддерживается кодек Рида-Соломона и декодер Витерби. Для построения оборудования на базе DM256 дополнительно необходим лишь ВЧ-трансмисмиттер и контроллер MAC-уровня. СБИС может использоваться как в БС, так и в абонентском оборудовании. Важно отметить, что DM256 совместима с требованиями грядущего расширения стандарта широкополосного доступа для мобильных приложений IEEE 802.16e.

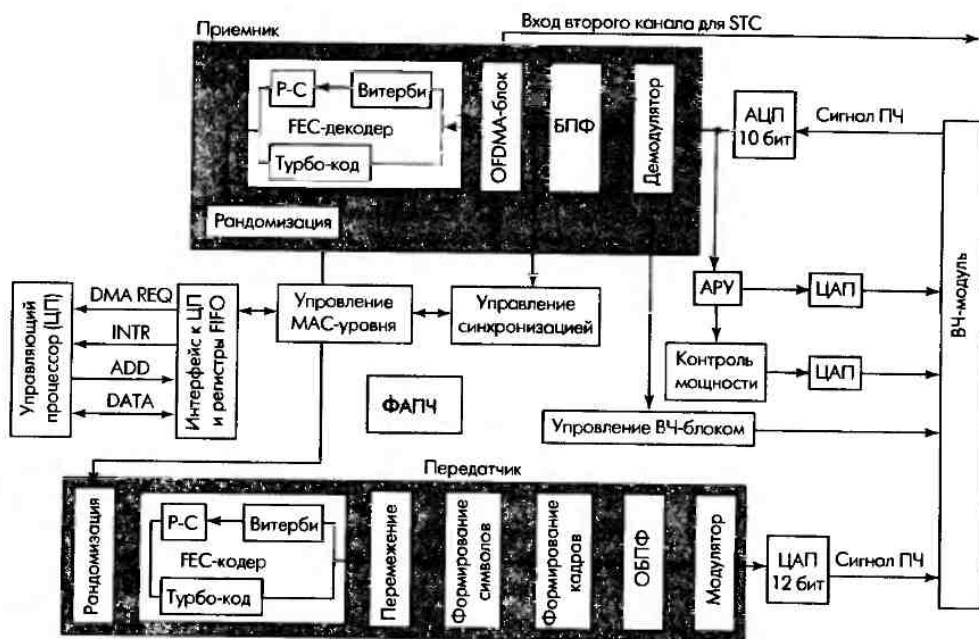


Рис. 8.26. Структура СБИС RN-2234 OFDMA-модема компании Runcom

Если компания Wavesat сосредоточилась на режиме OFDM, то израильская фирма Runcom Technologies (www.runcom.com) выпустила СБИС RN-2234 — модем с поддержкой режима OFDMA стандарта IEEE 802.16 (рис. 8.26). Данная СБИС реализует схемы модуляции QPSK, 16- и 64-QAM. Поддерживаются как турбокоды, так и кодек Рида-Соломона. Модем включает АЦП (10 бит) и ЦАП (12 бит). При ширине канала 14 МГц СБИС RN-2234 обеспечивает скорость передачи данных до 56 Мбит/с. На аппаратном уровне поддержана возможность работы с AAS, включая режим STC в нисходящем канале. Выпускается по КМОП-технологии 0,18 мкм в корпусе 304 LF BGA (19 × 19 × 1,5 мм). Напряжение питания ядра/периферии — 1,8/3,3 В. Максимальная потребляемая мощность — 2 Вт.

8.4.2 Особенности реализации аппаратуры стандарта IEEE 802.16

Одним из первых примеров реализации WiMax-совместимого оборудования стала система BreezeMAX 3500 компании Alvarion (Израиль) (рис. 8.27) [62, 63]. Она предназначена для работы в диапазоне 3400–3600 МГц. Для управления BreezeMAX 3500 служит система AlvariSTAR — платформа операторского класса, которая поддерживает полные функциональные возможности FCAPS (Fault, Configuration, Accounting, Performance and Security — защита от ошибок, конфигурация, статистика, производительность и уровень безопасности). БС BreezeMAX доступна в двух основных конфигурациях: базовая станция High Density для больших сетей и миниатюрная БС для сельских районов и зон с низкой плотностью населения.

Рис. 8.27. Система беспроводного доступа BreezeMAX 3500



Базовая станция High Density с PCI-шасси высотой 8U¹ подходит для установки в стандартные 19- или 22-дюймовые (ETSI) стойки. Шасси содержит модули сеговой обработки (NPU — Network Processing Unit), до шести устройств доступа в одиночном шасси и модули питания. Все модули допускают быструю замену и обладают высокой надежностью за счет схем избыточности. Устройство доступа состоит из внутреннего (IDU) и наружного (ODU) модулей. IDU-модуль содержит беспроводный модем и контроллер MAC-уровня. Он ответственен за установление связи в беспроводной сети и управление полосой пропускания. Каждый IDU поддерживает два физических канала шириной 3,5 или 1,75 МГц каждый. AU-ODU — это полнодуплексный, многоканальный радиомодуль с высокой выходной мощностью, который соединяется с внешней антенной. Он поддерживает полосу пропускания до 14 МГц и может обеспечивать многоканальность для увеличения пропускной способности за счет подключения дополнительных модулей IDU.

Миниатюрная базовая станция для сельских районов и областей с низкой плотностью населения состоит из автономного модуля, который соединяется с

¹Здесь U — это Unit, единица приращения размера конструкции по высоте в стандарте «Евромеханика», численно равная русскому вершку: 1U = 44,45 мм = 1 вершок.

наружным радиомодулем. Внутренний модуль высотой 1U вписывается в стандартные 19- или 22-дюймовые стойки, оснащен проводным интерфейсом Ethernet 10/100 Base-T и может питаться от сети переменного и постоянного тока.

С платформой BreezeMAX 3500 предлагаются несколько вариантов абонентского оборудования (CPE) для различных приложений со скоростью передачи до 12,7 Мбит/с в полосе 3,5 МГц. Абонентское оборудование BreezeMAX 3500 состоит из внутреннего (IDU) и наружного (ODU) модулей.

Наружный (ODU) модуль содержит все активные компоненты и плоскую интегрированную антенну с высоким усилением и соединяется с внутренним модулем посредством витой пары 5-й категории. Модуль ODU доступен в двух вариантах — с интегрированными антеннами (с вертикальной или горизонтальной поляризацией) или с разъемом для подключения внешней антенны. Различные типы абонентских устройств отличаются только внутренним модулем IDU.

Абонентское устройство с функциями моста данных используется для сопряжения беспроводных и проводных систем передачи информации и поддерживает до 512 MAC-адресов. Оно подключается через стандартный интерфейс IEEE 802.3 Ethernet 10/100-BaseT (RJ 45).

Абонентское устройство с функциями голосового шлюза обеспечивает интегрированные сервисы голоса и данных и доступно в двух моделях: с одним портом данных 10/100 Base-T и одним голосовым портом обычной телефонии (RJ-11); с четырьмя портами данных 10/100 Base-T и двумя голосовыми портами обычной телефонии (RJ-11). Устройство обладает VLAN маркировкой, поддерживает приоритезацию трафика по IEEE 802.1p и IP DiffServ, протоколы H.323 и SIP, голосовые сервисы 5-го класса (трехстороннюю конференцию, ожидание вызова, удержание вызова), интегрированное управление и многое другое.

Абонентское устройство с функциями сетевого шлюза представляет собой интегрированный широкополосный маршрутизатор с всесторонним IP-разделением и функциями безопасности. Оно производится в двух конфигурациях: как четырехпортовый 10/100 Base-T коммутатор и четырехпортовый 10/100 Base-T коммутатор с беспроводной точкой доступа IEEE 802.11 b/g.

В табл. 8.5 приводятся результирующие скорости оборудования при использовании всех поднесущих в режиме OFDM с определенным типом модуляции и кодирования [45, 234].

Ожидается появление портативных абонентских устройств и устройств на основе СБИС производства Intel.

8.5 Будущее широкополосного беспроводного доступа по стандартам IEEE 802.16

Итак, конец 2004 года считается началом выхода на рынок оборудования, совместимого с WiMax, различных компаний — Airspan, Alvarion, Aperto, Proxim и др. на собственной элементной базе для различных диапазонов частот: 2,5–2,7, 3,4–3,6, 5,725–5,85 ГГц и, возможно, некоторых других. К середине 2005 года ожидается начало серийного производства на основе чипсетов компании Intel и, возможно, других производителей чипсетов. При удачном развитии событий и

достаточной конкуренции на рынке уже к началу 2006 года абонентское оборудование значительно подешевеет.

Таблица 8.5. Параметры передачи в системе BreezeMAX 3500

Модуляция, скорость кодирования	3.5 МГц канал		1.75 МГц канал	
	Битовая скорость, Мбит/с	Чувствитель- ность, дБм	Битовая скорость, Мбит/с	Чувствитель- ность, дБм
BFSK 1/2	1,41	-100,0	0,71	-103,0
BFSK 3/4	2,12	-98,0	1,06	-101,0
QPSK 1/2	2,82	-97,0	1,41	-100,0
QPSK 3/4	4,23	-94,0	2,12	-97,0
QAM16 1/2	5,64	-91,0	2,82	-94,0
QAM16 3/4	8,47	-88,0	4,24	-91,0
QAM64 2/3	11,29	-83,0	5,65	-86,0
QAM64 3/4	12,71	-82,0	6,35	-85,0

По-видимому, в ближайшее время следует ожидать появления оборудования с тремя возможными профилями:

- 1) рабочий диапазон 2,5–2,7 ГГц, ширина канала — не более 6 ГГц, временное дуплексирование;
- 2) рабочий диапазон 3,4–3,6 ГГц (возможны небольшие отклонения), ширина канала — 1,75; 3,5; 7 ГГц и более, с частотным дуплексированием;
- 3) рабочий диапазон 5,725–5,850 ГГц (возможны небольшие отклонения) с шириной канала 5 ГГц.

С появлением к 2006 году стандарта IEEE 802.16e, поддерживающего мобильность, ожидается новый всплеск интереса к этому бизнесу. По-видимому, технология WiMax будет претендовать на отдельный крупный вид телекоммуникационного бизнеса, занимая нишу где-то между мобильными сотовыми сетями, WiFi сетями и проводными сетями, соответственно отбирая клиентов у всех этих сетей. Насколько успешным будет этот процесс, покажет время.

Сейчас принято выделять четыре градации абонентских устройств по степени «подвижности»:

- 1) неподвижные абонентские устройства (fixed wireless), как правило, с внешним (outdoor) и внутренним (indoor) блоками и узконаправленными антеннами; эти устройства требуют профессиональной установки;
- 2) самоинсталируемые (self installed) абонентские устройства, состоящие только из внутреннего блока и имеющие антенну с широкой диаграммой направленности; их абонент может устанавливать самостоятельно;

- 3) портативные (portable) абонентские устройства, отличающиеся от самонсталируемых только тем, что могут работать при передвижении абонента с места на место со скоростью до 5 км/час;
- 4) мобильные (mobile) абонентские устройства, обеспечивающие работу при передвижении абонента со скоростью до 120 км/час.

Поскольку мобильная версия стандарта еще не принята, нельзя определенно заявлять, в каком диапазоне частот будет обеспечена мобильность. Однако можно утверждать с уверенностью, что мобильности не будет в 3- и 5-ГГц диапазонах. По оценкам аналитиков, наиболее вероятными с точки зрения мобильности могут оказаться диапазоны 2,7–3,5 ГГц и 700 МГц.

В России с момента начала развития беспроводного доступа отечественные компании пытались активно участвовать в производстве оборудования для беспроводных сетей. К сожалению, в основном производство касалось вспомогательных аксессуаров — кабелей, разъемов, усилителей, грозозащитников, антенн и т.д. Однако не правильно было бы ограничивать роль промышленности России только производством радиоаксессуаров. Огромный научно-технический потенциал России не утерян и до сих пор [185, 199]. Об этом свидетельствует тот факт, что многие мировые лидеры в области технологии беспроводного оборудования используют коллективы российских разработчиков. Не следует также забывать, что научная база для практической реализации OFDM-систем впервые была описана российскими учеными [46]. Сказанное выше позволяет надеяться, что Россия в будущем не только будет полигоном для быстрого и широкомасштабного внедрения WiMax-совместимых систем, но и займет достойное место в мировом разделении рынка по производству таких систем. Подтверждением этого является появление на рынке отечественного радиомаршрутизатора «РАПИРА» [18], описание которого приводится в главе 9, и разработка ряда отечественных устройств для комплектации беспроводных сетей [15, 17, 19].

ГЛАВА 9

Архитектура и технические средства беспроводных региональных сетей

9.1 Архитектура беспроводных региональных сетей с фиксированной топологией

В данном разделе рассматриваются способы и средства построения фиксированных беспроводных широкополосных сетей передачи данных. Фиксированная топология таких сетей означает отсутствие мобильных станций и используется в основном для построения корпоративных сетей и сетей операторов связи городского и регионального масштаба. Высокая распространенность таких сетей обусловлена слабой развитостью кабельной среды в городах и обширных сельских районах Российской Федерации с небольшой плотностью населения.

В общем случае рассматриваемая сеть имеет сотовую структуру и состоит из радиосот — базовых станций и абонентских станций. Базовые станции образуют единую инфраструктуру и соединены между собой посредством опорной сети, которая может быть построена с помощью как проводных, так и беспроводных средств связи. Типичным примером построения беспроводной широкополосной сети является сеть RadioNet, разработанная и реализованная в 1996 году. Институтом проблем передачи информации Российской академии наук (ИППИ РАН) для подключения организаций науки и образования г. Москвы к сети Интернет. Фрагмент этой сети представлен на рис. 9.1.

Сеть функционирует в диапазоне 2,4 ГГц, обеспечивая высокоскоростной и надежный доступ в Интернет для 5000 ПЭВМ.

Базовая станция состоит из:

- антенно-фидерной системы, в которую могут входить антенна, высокочастотный кабель снижения и усилитель;
- приемопередающего оборудования (радиомоста, радиомаршрутизатора);

- коммутационного и маршрутизирующего оборудования;
- точки доступа к опорной сети.

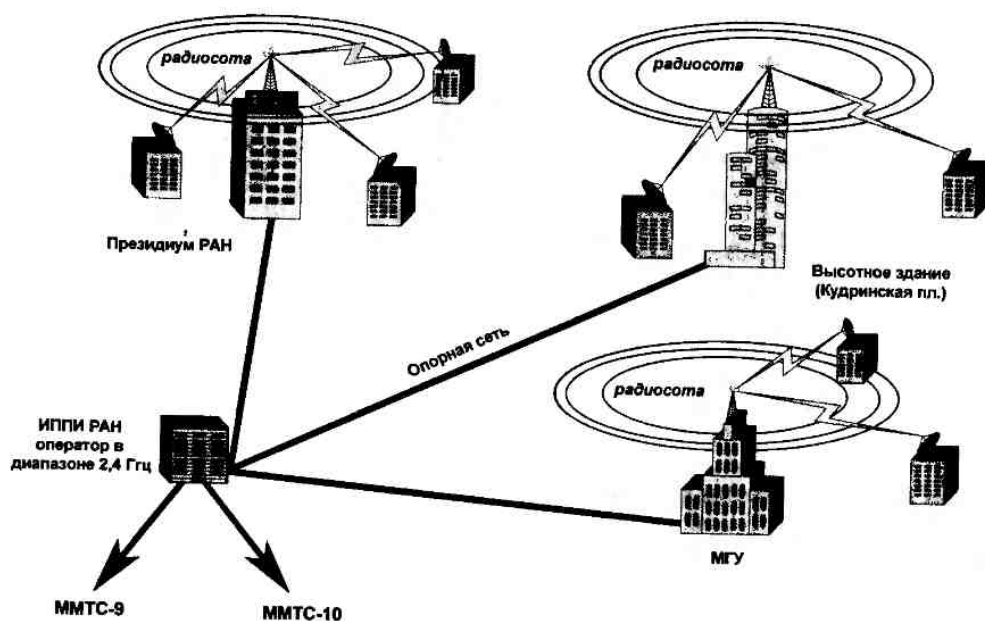


Рис. 9.1. Беспроводная широкополосная сеть RadioNet

В целях обеспечения прямой видимости между базовой и абонентскими станциями базовые станции размещаются либо на высотных зданиях, либо на телекоммуникационных вышках. При этом активное оборудование часто устанавливается в отапливаемых и кондиционируемых помещениях, а антенные системы монтируются максимально высоко на крыше здания или пике вышки (рис. 9.2). Состав базовой станции может варьироваться в зависимости от исполнения приемопередающего оборудования, которое может быть:

- комнатным: оборудование устанавливается в отапливаемом помещении. Используется высокочастотный кабель снижения, причем его длина может достигать 60–70 м. В связи с этим для компенсации потерь в кабеле могут использоваться высокочастотные усилители, которые устанавливаются на антенной мачте рядом с антенной (варианты А и В на рис. 9.2);
- всепогодным: в этом случае используется низкочастотный кабель снижения, по которому передаются цифровые данные и подается электропитание к оборудованию. Оборудование монтируется на антенной мачте в непосредственной близости к антенне, при этом длина высокочастотного кабеля обычно не превышает 10 м. Возможна установка высокочастотного усилителя для увеличения радиуса охвата базовой станции (вариант Б на рис. 9.2);

- всепогодным со встроенной антенной: также используется низкочастотный кабель снижения. Установка высокочастотного усилителя невозможна (вариант Г на рис. 9.2).

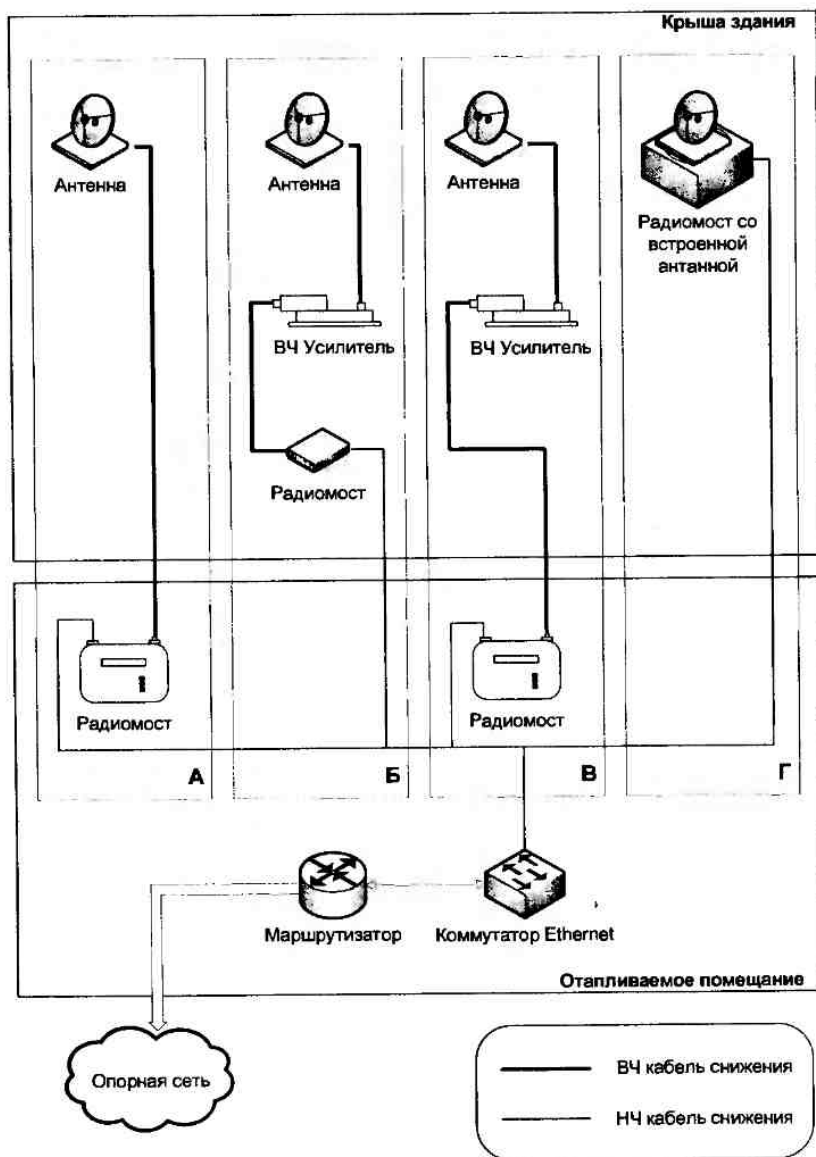


Рис. 9.2. Состав оборудования базовой станции

Базовые станции размещаются таким образом, чтобы обеспечить покрытие определенной зоны обслуживания. При этом могут применяться различные типы антенн — от простейших штыревых с круговой диаграммой направленности

до сложных систем секторных антенн, позволяющих изменять диаграмму направленности излучения в зависимости от потребностей (см. раздел 9.4). Каждое приемопередающее устройство (радиомост, радиомаршрутизатор), работающее в режиме базовой станции, позволяет подключить 15–20 абонентов. В непосредственной близости на одной крыше или вышке при соблюдении определенных условий по развязке поляризации и частотному разнесению можно установить до шести комплектов оборудования базовой станции, что позволит увеличить количество подключаемых абонентов к одной базовой станции до 90–100.

Абонентская станция в общем случае также состоит из:

- антенно-фидерной системы, в которую могут входить антенна, высокочастотный кабель снижения и усилитель;
- приемопередающего оборудования (радиомоста, радиомаршрутизатора);
- коммутационного и маршрутизирующего оборудования.

Антенна устанавливается в доступное для абонента место, в котором обеспечивается прямая видимость с базовой станцией. В условиях высоких городских радиопомех, а также при достаточно большой удаленности от базовой станции (больше 5 км) необходима установка высокочастотного усилителя. Идеальным исполнением клиентского приемопередающего оборудования является радиомаршрутизатор со всепогодным исполнением, так как это значительно удешевляет стоимость комплекта вследствие интеграции в одном устройстве функций радиомоста и маршрутизатора локальной сети абонента, а также вследствие использования низкочастотного кабеля снижения, что снижает трудоемкость его монтажа (см. раздел 9.2).

9.2 Логическая и аппаратная структура радиомаршрутизатора «РАПИРА»

За годы эксплуатации сети Radionet в ИППИ РАН был накоплен большой практический опыт в области строительства и эксплуатации беспроводных широкополосных сетей, детально изучен спектр сопутствующих проблем. При этом в образцах оборудования различных производителей были выявлены конструктивные и протокольные недостатки, которых удалось избежать в собственной разработке РЭС «РАПИРА».

В настоящее время основными производителями и поставщиками на российский рынок оборудования беспроводной передачи данных являются зарубежные компании. К числу достоинств их продукции следует отнести достаточно высокое качество, но применительно к российским условиям это оборудование связано с неприятными ограничениями, сильно сдерживающими его применение отечественными операторами. В числе таких недостатков — высокая стоимость, низкая приспособленность устройств к местным особенностям климата и распределения радиочастотного спектра. К тому же необходимо заметить, что оборудование западных компаний ориентировано в основном на сектор SOHO (Small Office/Home

Office), в то время как отечественные разработки нацелены на построение «последней мили», что весьма актуально в России с ее обширными территориями и слабо развитой телекоммуникационной инфраструктурой в сельских регионах [16].

Все упомянутые факторы создали реальную почву для разработки и серийного производства новой отечественной радиоэлектронной системы «РАПИРА» (сокращенно РЭС «РАПИРА»), которая:

- доступна по цене для российских операторов,
- максимально подготовлена для применения в условиях российского законодательства в области связи и распределения радиочастотного спектра.

В данном разделе мы последовательно будем рассматривать проблемы, которые возникают при применении беспроводных широкополосных устройств, а также опишем решения, нашедшие воплощение в новой разработке.

9.2.1 Аппаратная платформа

РЭС «РАПИРА» — высокопроизводительное устройство, предназначенное для создания территориальнораспределенных широкополосных сетей беспроводного абонентского доступа к ресурсам Интернета, телефонии и других сетей общего пользования, а также создания корпоративных и ведомственных сетей с интеграцией голоса, видео, телеметрии и т.д. На физическом уровне РЭС «РАПИРА» совместима со всеми устройствами стандартов IEEE 802.11 a/b/g. Разработанное устройство обладает полным набором характеристик, присущих современному маршрутизатору, а также рядом специальных функций, необходимых при использовании устройства в составе широкополосных радиосетей. Как правило, радиомаршрутизатор является точкой доступа широкополосной радиосети, которая, в свою очередь, является точкой выхода во внешнюю сеть, но, наряду с этим, «РАПИРА» может выполнять функции любого элемента широкополосной радиосети: базовой станции (БС), абонентской станции (АС), а также базовой станции с ретрансляцией (WDS Repeater). В модификации маршрутизатора с двумя радиointерфейсами каждый из них может быть независимо сконфигурирован, при этом устройство может выполнять сразу две функции, например БС и АС.

РЭС «РАПИРА» представляет собой аппаратную платформу (рис. 9.3), имеющую проводные и беспроводные интерфейсы, под управлением специализированной операционной системы, реализующей функции маршрутизатора.

Основными компонентами аппаратной платформы являются материнская плата и плата (платы) радиointерфейсов. Материнская плата содержит встроенные: процессор с тактовой частотой 266 МГц, оперативную память типа SDRAM в размере 64 Мбайт, контроллер (контроллеры) Ethernet, разъемы miniPCI и разъем для подключения Compact Flash. Радиointерфейсы реализованы на базе беспроводных miniPCI адаптеров, выполненных на чипсетах Atheros, и имеют следующие возможности:

- поддержка протоколов IEEE 802.11 a/b/g,
- поддержка турборежима, позволяющего достигать скорости 108 Мбит/с в канале передачи данных,

- аппаратное шифрование для Wi-Fi Protected Access (WPA) и стандарта IEEE 802.11i с использованием Advanced Encryption Standard (AES), Temporal Key Integrity Protocol (TKIP) и Wired Equivalent Privacy (WEP)

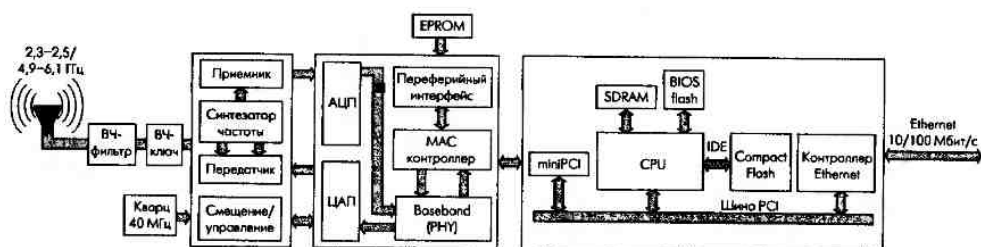


Рис. 9.3. Блок-схема аппаратной платформы РЭС «РАПИРА»

9.2.2 Проблема частотного ресурса

Любой оператор, решивший строить свою сеть на базе радиосистем, неизбежно сталкивается с так называемой «проблемой частотного ресурса».

Зарубежное оборудование беспроводной широкополосной связи работает в привычных нам диапазонах 2,4–2,5, 5,15–5,35, 5,725–5,850 ГГц, поскольку указанные частоты являются нелицензируемыми в большинстве развитых стран, но не в России, где соответствующие полосы относятся к классу ПР (правительственных) и могут быть выделены операторам лишь на вторичной основе.

Полоса 2,4–2,5 ГГц в большинстве регионов страны полностью исчерпана и чрезвычайно зашумлена. Диапазон стал одним из первых, для которого с 1995 года стали производиться образцы беспроводного оборудования передачи данных. С тех пор в нашу страну не прекращается колоссальный поток «серых» поставок оборудования, которое находит широкое применение у нелегальных пользователей. В результате этот диапазон стал практически непригоден для коммерческого применения и перестает использоваться легальными операторами связи.

Полосы 5,15–5,35 и 5,725–5,850 ГГц в ближайшее время повторят участь диапазона 2,4 ГГц, поскольку активно осваиваются операторами, покидающими участок 2,4 ГГц. При этом объемы «серых» поставок зарубежного оборудования растут высокими темпами. Кроме того, в большинстве регионов России данный участок спектра уже закрепили за собой различные структуры с перспективой выгодно реализовать его в будущем. Таким образом, легальное оказание услуг в этом участке спектра сильно затруднено и доступно далеко не везде.

Полоса 5,65–5,725 в ходе проведенных исследований была выделена ГКРЧ для работы беспроводных широкополосных устройств на вторичной основе. Указанный участок спектра перекрывается с принятым в европейском сообществе 5,47–5,725. В силу ряда причин зарубежное оборудование на этот участок спектра распространено слабее, чем на все упомянутые выше.

Ресурс полосы 3,2–3,8 ГГц, на которую ориентирован стандарт беспроводного оборудования Wi-MAX 802.16a, в значительной степени исчерпан радиорелейными и прочими узкополосными системами, установленными в предыдущие годы,

что привело к наложению ограничений на работу в этом спектре, который доступен операторам лишь частично.

Решающее преимущество РЭС «РАПИРА» заключается в исключительно широком диапазоне частот, в котором функционирует указанное оборудование (рис. 9.4). Серийные модификации РЭС «РАПИРА» работают в диапазонах частот от 2,3 до 2,5 ГГц и от 4,9 до 6,1 ГГц.

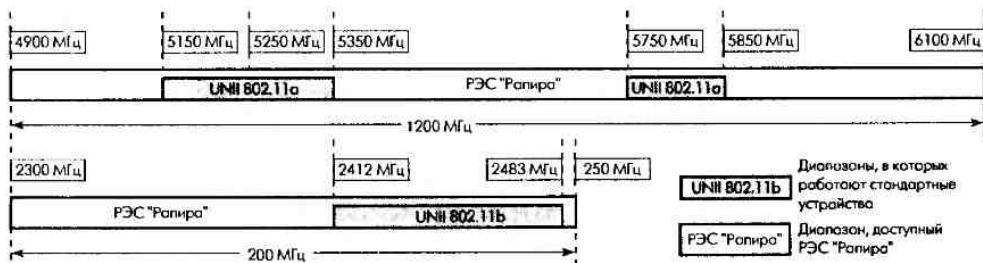


Рис. 9.4. Диапазон рабочих частот РЭС «РАПИРА»

В большинстве промышленных и густонаселенных регионов России получение частотных разрешений и работа в «нестандартных» диапазонах 2,3–2,4 ГГц, а также 5,65–5,75 и 5,85–6,1 ГГц является самым простым и разумным способом стать легальным оператором и работать в чистом эфире.

По всем прогнозам диапазоны 2,3 и 6 ГГц будут оставаться чистыми в течение длительного времени вследствие отсутствия «серого» зарубежного оборудования на этот диапазон частот.

9.2.3 Логическая схема абонентского тракта

РЭС «РАПИРА» реализует функциональность нескольких устройств (рис. 9.5):

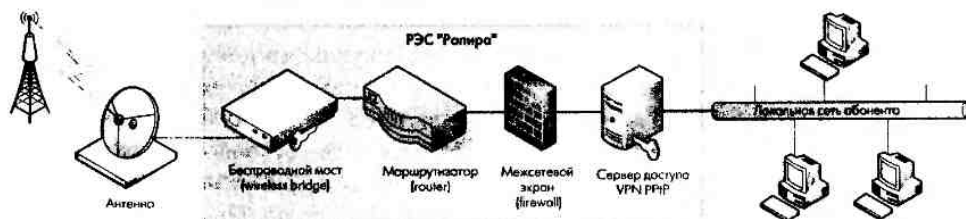


Рис. 9.5. Логическая схема абонентского тракта

- беспроводного модема с развитыми средствами наблюдения за трафиком,
- маршрутизатора с поддержкой протоколов динамической маршрутизации *rip*, *ospf* и *bgr*, а также *QoS* на уровне *TCP/IP*,
- межсетевого экрана с возможностью анализа и фильтрации трафика вплоть до уровня 7 модели *OSI*,

- сервера доступа, обеспечивающего создание туннелей VPN, PPTP, шифрованных MPPE и авторизацию пользователей по протоколам PAP, CHAP, MS-CHAP.

Подавляющее большинство представленных на рынке беспроводных устройств, предназначенных для построения городских и территориальнораспределенных беспроводных сетей, представляют собой простые мосты, т.е. устройства, работающие на канальном (MAC) уровне модели OSI и осуществляющие передачу кадров Ethernet из проводной среды в беспроводную и наоборот.

Теперь предположим, что оператор имеет беспроводную сеть с базовой станцией БС, клиентами которой являются мосты АС, которые обеспечивают подключение локальных сетей корпоративных абонентов. Какие проблемы заложены в такой схеме подключения?

1. Проблема безопасности. Поскольку беспроводные мосты БС и АС прозрачны для Ethernet трафика, созданная сеть фактически оказывается связанной совокупностью локальных сетей отдельных абонентов, т.е. одной большой локальной сетью. Такая ситуация не удовлетворяет заказчиков, заботящихся о своей сетевой безопасности. В качестве решения проблемы некоторые модели точек доступа предлагают режим блокирования обмена пакетами между различными клиентами АС, ассоциированными с одной БС, что, однако, делает невозможным обмен трафиком между клиентами в тех случаях, когда это действительно требуется, и не избавляет от описанных ниже проблем.

2. Проблема широковещательной нагрузки. Функционирование целого ряда протоколов (ARP, NETBIOS и др.) сопровождается регулярной рассылкой станциями широковещательных пакетов. В крупных Ethernet сетях под управлением Windows доля широковещательного трафика может составлять существенный процент общей нагрузки на сеть. Поскольку сеть, построенная на мостах, является одной большой локальной сетью, широковещательные пакеты всех присоединенных клиентских сетей будут циркулировать в ней, существенно увеличивая нагрузку на каналы.

3. Проблема управляемости. Предположим, что локальная сеть одного из абонентов стала источником паразитного, например вирусного, трафика, который занимает полосу пропускания беспроводных каналов, затрудняя доступ абонентов к ресурсам сети и Интернета. Предположим, что эта неприятность обнаружена на главном маршрутизаторе. Как оперативно устранить ее, отфильтровав трафик на стороне клиента, если отсутствуют рычаги контроля на уровне TCP/IP?

4. Проблема отсутствия QoS. Современные сетевые приложения реального времени, такие как IP-телефония, получившие широкое распространение у корпоративных и частных клиентов, предъявляют весьма жесткие требования к качеству обслуживания трафика. Совместная передача по беспроводному каналу трафика IP-телефонии и Best-effort (например, HTTP или FTP) даже при минимальной загруженности канала начнет сильно сказываться на качестве голосового соединения, которое будет страдать из-за увеличения задержки пакетов в канале и сильной вариабельности этой задержки в зависимости от нагрузки на канал (jitter).

Тем не менее такая схема работы возможна, если беспроводное устройство реализует функции QoS хотя бы на уровне IP. В этом случае встроенный маршрутизатор будет непрерывно анализировать пакеты данных, попадающие в его очереди на отправку, проводя так называемые «классификацию» и «приоритезацию». При этом пакеты голосового трафика «пойдут всегда первыми».

При правильной схеме подключения локальная сеть заказчика должна быть отделена от беспроводной сети маршрутизатором и сетевым экраном, что, разумеется, удорожает стоимость инсталляции для каждого из клиентов на несколько сотен долларов. Простейшие недорогие SOHO-маршрутизаторы обладают весьма ограниченной производительностью, их полоса пропускания редко превышает 2 Мбит/с, эти устройства способны лишь отчасти справиться с проблемами 1 и 2.

Как уже отмечалось, радиомаршрутизаторы «РАПИРА» сочетают в себе функциональность всех необходимых компонентов и способны не только решить все указанные проблемы, но и помочь сэкономить на покупке дополнительных устройств — фаерволов, маршрутизаторов, серверов доступа. В тех случаях, когда требуется организовать ограниченный доступ в сеть с авторизацией по паролю и подсчетом трафика для клиентов локальной сети (офиса, жилого здания), устройство «РАПИРА» вполне способно справиться с функцией сервера доступа RADIUS, взаимодействующего со стандартным сервером RADIUS.

Кроме того, унификация встроенных абонентских маршрутизаторов в сочетании с развитыми возможностями дистанционного управления этими устройствами (SSH, SNMP, WEB HTTP) дает возможность производить мониторинг сетей заказчиков непосредственно из сетевого операционного центра беспроводного оператора. Таким образом, оператор может создавать добавленную стоимость для своих услуг передачи данных и аренды каналов.

9.2.4 Обеспечение информационной безопасности

Информационная безопасность всегда вызывала резонное беспокойство у существующих и потенциальных пользователей беспроводных сетей. Это не удивительно, поскольку в отличие от кабеля, зарытого глубоко под землей, радиозфир доступен для «прослушивания».

Круг потенциальных проблем безопасности при эксплуатации сети сводится к трем основным:

- 1) прослушивание эфира в целях хищения важной информации клиентов;
- 2) прослушивание эфира и подмена «на лету» передаваемых данных (forgery);
- 3) несанкционированное подключение в целях кражи трафика Интернета.

Алгоритм WEP, который должен был помочь решить проблему беспроводной сетевой безопасности и получил широкое распространение в серийных беспроводных устройствах стандарта 802.11, оказался уязвимым для различного рода атак и не обеспечивает достаточной надежности.

На смену WEP приходят стандарты IEEE 802.11i и WPA от Wi-Fi Alliance.

Структуру защитной технологии WPA можно наглядно выразить с помощью следующей формулы: $WPA = 802.1X + EAP + TKIP + MIC + (RADIUS * X)If$

PSK, $X = 0$; ELSE $X = 1$. Из этой формулы видно, что в стандарте WPA предусмотрено использование известных специалистам защитных протоколов 802.1x, EAP, TKIP и RADIUS. Механизм аутентификации пользователей основан на протоколах 802.1x (разработан для проводных сетей) и EAP (Extensible Authentication Protocol). Последний позволяет сетевому администратору задействовать множество алгоритмов аутентификации пользователей посредством сервера RADIUS.

В технологии WPA функции обеспечения конфиденциальности и целостности данных базируются на протоколе TKIP, который в отличие от протокола WEP использует более эффективный механизм управления ключами, но тот же самый алгоритм RC4 для шифрования данных. Согласно протоколу TKIP сетевые устройства работают с 48-битовым вектором инициализации (в отличие от 24-битового вектора инициализации протокола WEP) и реализуют правила изменения последовательности его битов, что исключает повторное использование ключей и осуществление replay-атак. В протоколе TKIP предусмотрены генерация нового ключа для каждого передаваемого пакета и улучшенный контроль целостности сообщений с помощью криптографической контрольной суммы MIC (Message Integrity Code), препятствующей изменению хакером содержимого передаваемых пакетов (forgery-атака).

Система сетевой безопасности стандарта WPA работает в двух режимах: PSK (Preshared Key) и Enterprise (корпоративный). Для развертывания системы, работающей в режиме PSK, необходим разделяемый пароль. Такую систему несложно устанавливать, но она защищает беспроводную сеть не столь надежно, как это делает система, функционирующая в режиме Enterprise с иерархией динамических ключей. Хотя протокол TKIP работает с тем же самым блочным шифром RC4, который предусмотрен спецификацией протокола WEP, технология WPA защищает данные надежнее последнего.

Между тем определенный по умолчанию в стандарте IEEE 802.11i механизм обеспечения конфиденциальности данных основан на блочном шифре стандарта AES. Используемый его защитный протокол получил название Counter-Mode CBC MAC Protocol, или CCMP. Для этого протокола AES играет ту же роль, что и RC4 для протокола TKIP. Основное различие между протоколами CCMP и TKIP проявляется на нижних уровнях модели OSI, где происходят шифрование и расшифровка передаваемых данных: TKIP использует четыре временных ключа шифрования, тогда как AES — только три. Оба протокола работают с одним и тем же механизмом управления ключами.

Чтобы корпоративные точки доступа работали в системе сетевой безопасности стандарта WPA или 802.11i, они должны поддерживать аутентификацию пользователей по протоколу RADIUS и реализовывать предусмотренный стандартом метод шифрования — TKIP или AES. Еще одно требование — быстро осуществлять повторную аутентификацию пользователей после разрыва соединения с сетью. Это особенно важно для нормального функционирования приложений, работающих в реальном масштабе времени (например, средств передачи речи по беспроводным сетям).

Если сервер RADIUS, применяемый для контроля доступа пользователей проводной сети, поддерживает нужные методы аутентификации EAP, его можно задействовать и для аутентификации пользователей беспроводной сети. Сервер

WLAN RADIUS работает следующим образом: сначала он проверяет аутентифицирующую информацию пользователя (на соответствие содержимому своей базы данных об их идентификаторах и паролях) или его цифровой сертификат, а затем активизирует динамическую генерацию ключей шифрования точкой доступа и клиентской системой для каждого сеанса связи. Для работы технологии WPA требуется механизм EAP-TLS (Transport Layer Security), тогда как в стандарте IEEE 802.11i применение каких-либо конкретных методов аутентификации EAP не оговаривается. Выбор метода аутентификации EAP определяется спецификой работы клиентских приложений и архитектурой сети.

РЭС «РАПИРА» реализует все описанные современные средства обеспечения безопасности в беспроводных сетях, в том числе WEP (64,128,154 бита), WPA-PSK, WPA-EAP с шифрованием TKIP и AES.

9.2.5 Температурная стабилизация в РЭС «РАПИРА»

подавляющее большинство беспроводного сетевого оборудования, представленного на рынке, имеет комнатное, офисное реже — стойное исполнение. Известно, что практическое применение такого оборудования сопряжено с серьезными проблемами:

- высокочастотный кабель снижения (фидер), соединяющий собственно беспроводный модем, мост или маршрутизатор с антенной, вносит весьма существенное затухание в принимаемый и передаваемый сигнал. Особенно этот аспект становится актуальным с переходом на высокочастотные диапазоны 5–6 ГГц;
- в непосредственной близости от площадки, где планируется размещение антенн, часто отсутствует подходящее помещение для размещения оборудования (отапливаемое и достаточно сухое);
- длинный кабель снижения (фидер) становится паразитной антенной, улавливающей мощные импульсы электромагнитного излучения, сопровождающего грозовые явления в атмосфере. Наведенная ЭДС вполне способна вывести из строя чувствительное беспроводное оборудование;
- высокочастотный кабель с низкими потерями обладает достаточно высокой ценой, не менее 3–5 долл. за метр, что удорожает установку;
- монтаж и крепление длинного фидера требуют специальных усилий, знания технологии и часто сопряжены с необходимостью проведения дополнительных небезопасных работ на высоте, что в конечном счете также сильно увеличивает стоимость установки.

Предположим, что антенна расположена на мачте высотой 20 м, мост расположен в помещении у ее подножия. Связь между антенной и мостом осуществляется ВЧ кабелем марки 8D-FB длиной 30 м. На частоте 6 ГГц затухание сигнала в этом отрезке кабеля составит порядка 20 дБ, что сопоставимо с усилением хорошей направленной антенны. Таким образом, мощность сигнала уменьшается

в 100 раз. Кроме этого, стоимость такого фидера составляет порядка 100 долл., не считая стоимости монтажа. Компенсирующий усилитель, который мог бы помочь решить данную проблему, удорожает инсталляцию еще на 700 долл.

Специально для решения указанных проблем РЭС «РАПИРА» выпускается в защищенном всепогодном исполнении, позволяющем размещать его прямо на улице под открытым небом в непосредственной близости от антенны. При этом радиомаршрутизатор соединяется с антенной кабелем длиной не более 1 м, что делает потери в фидере пренебрежимо малыми.

Климат нашей страны и диапазон уличных температур варьируются в широких пределах в зависимости от региона и времени года. В новом радиомаршрутизаторе удалось решить непростую задачу, связанную с надежной герметизацией электронной начинки оборудования, термостатированием электроники в широком диапазоне температур. Комплектуемые, использованные при производстве модема, специально отбирались исходя из широкого диапазона допустимых температур эксплуатации. Добротный корпус выполнен из толстого литого алюминия, снабжен ребрами

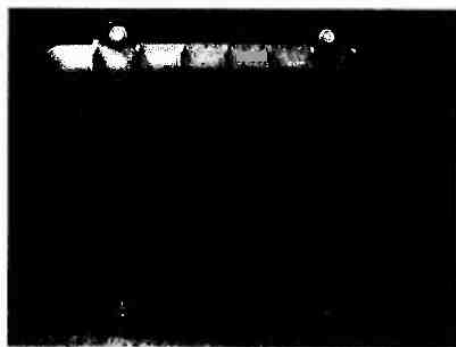


Рис. 9.6. Внешний вид РЭС «РАПИРА»

жесткости и не боится ни дождя, ни града (рис. 9.6). Герметизация соответствует стандарту IP67, т. е. устройство способно выдержать погружение в воду.

9.2.6 Использование технологии PoE (Power over Ethernet)

Подача питающего напряжения к оборудованию, расположенному на площадке — мачте или крыше, представляет собой особую проблему, поскольку в целях обеспечения безопасности прокладка дополнительного кабеля 220 В в уличных условиях должна выполняться специальным влагозащитным кабелем с применением влагозащитных разъемов. Все это неминуемо усложняет и удорожает инсталляцию.

РЭС «РАПИРА» использует перспективную технологию PoE (Power over Ethernet) IEEE 802.3af для подачи безопасного напряжения питания 18 В непосредственно по кабелю Ethernet «витая пара». За счет применения этой технологии инсталляция оборудования существенно упрощается и удешевляется. Единственный кабель, который требуется проложить к оборудованию, размещенному на мачте или крыше, — экранированная витая пара FTP или STP Cat.5.

Обеспечение бесперебойной работы сетевого оборудования — еще одна важная проблема, эффективное и недорогое решение которой удастся далеко не всегда. Привычные источники бесперебойного питания (ИБП) переменного тока не могут обеспечить длительного многочасового питания оборудования при работе от батарей, что связано с невысоким КПД преобразования энергии низковольтных аккумуляторов постоянного тока в переменное напряжение 220 В. Стоимость

ИБП, допускающего подключение внешней батареи большой емкости, может составить сумму, сопоставимую со стоимостью самого беспроводного оборудования. Кроме того, питаемое оборудование может дать сбой из-за перерыва в подаче питания во время переключения на батареи.

Существует специальный класс недорогих ИБП постоянного тока, обеспечивающих практически 100% КПД за счет отсутствия ступени преобразования напряжения. При перебое с питанием нагрузка таких ИБП получает постоянное напряжение 12–48 В непосредственно от аккумуляторов. Время переключения такого источника на батареи стремится к нулю.

Система «РАПИРА» допускает питание от нестабилизированных источников в широком диапазоне напряжений и специально разрабатывалась для работы с ИБП постоянного тока. Низкое энергопотребление РЭС «РАПИРА» в сочетании с высочайшим КПД источника делает возможной ее автономную работу в течение суток и более, причем для этого достаточно применения ИБП постоянного тока стоимостью менее 100 долл.

9.3 Фидерные линии и элементы СВЧ-тракта

9.3.1 Основные понятия

Фидерные линии или линии передачи (ЛП) – это радиотехнические устройства, предназначенные для передачи энергии электромагнитных волн [36, 79]. В настоящее время используются следующие типы ЛП:

- двухпроводные открытые и экранированные;
- коаксиальные;
- прямоугольные и круглые волноводы;
- полосковые линии и волноводы.

Схематически всякую фидерную линию можно изобразить в виде двухпроводной линии, к которой с одного конца присоединен источник э.д.с., а с другого — нагрузка, например антенна. Физическими параметрами фидерных линий являются: волновое сопротивление, коэффициент затухания и коэффициент скорости, равный отношению скорости света в вакууме к фазовой скорости в фидере. Эти параметры зависят от типа линии (двухпроводная, коаксиальная, волноводная и т. п.), от ее размеров, типа (порядка) волны, частоты электромагнитных колебаний и того материала, из которого сделана линия.

Процесс передачи электромагнитной энергии вдоль линии независимо от ее типа всегда носит волновой характер. От генератора к нагрузке распространяется падающая волна, причем в длинной линии с потерями амплитуда падающей волны убывает в сторону нагрузки. Если нагрузка не согласована с линией, то часть мощности отражается, и возникает отраженная волна.

Комплексный коэффициент отражения ρ характеризует соотношение амплитуд падающей и отраженной волн, а также сдвиг фаз этих колебаний в рассматриваемой точке. Значение модуля ρ лежит в пределах от 0 до 1, так как амплитуда отраженной волны не может превышать по величине амплитуду падающей.

Если нагрузка согласована с ЛП и полностью поглощает падающую на нее мощность, то отраженной волны не будет ($p = 0$), и в линии установится *режим бегущих волн*. Амплитуды напряжения (тока) вдоль линии в этом случае остаются постоянными, меняется лишь фаза колебаний вдоль фидера.

Если нагрузка не рассеивает активную мощность и полностью отражает падающую волну, то $|p| = 1$, т.е. амплитуда отраженной волны равна амплитуде падающей. В точках, где фазы напряжений падающей и отраженной волн совпадают, коэффициент отражения p является вещественной величиной. Результирующее напряжение и ток в случае полного отражения (при $|p| = 1$) представляют картину *стоячих волн* с узлами и пучностями.

Если нагрузка на конце линии часть мощности поглощает, а часть отражает, то значение модуля коэффициента отражения находится в пределах $0 < |p| < 1$. Получившийся *смешанный режим* является промежуточным между режимом бегущих и стоячих волн. Сечения максимумов и минимумов в распределении амплитуд напряжений и токов в линии называются характерными сечениями. В сечениях, где амплитуда напряжения максимальная, амплитуда тока будет минимальной, и наоборот. Расстояние между двумя соседними максимумами (минимумами) $\Delta z = \lambda/2$. Точки минимума и максимума напряжения смещены друг относительно друга на постоянное расстояние $\lambda/4$.

Во всех характерных сечениях ЛП входное сопротивление линии является активным, тогда как в прочих сечениях оно имеет комплексный характер.

Если нагрузка линии не равна волновому сопротивлению, в линии, как указывалось, наряду с падающей появляется отраженная волна. Чем больше $|p|$, тем больше амплитуда отраженной волны и тем ближе режим в линии к режиму стоячих волн. За счет отраженной волны увеличиваются потери, появляются перенапряжения, что снижает электрическую прочность фидера, усиливается зависимость входного сопротивления от частоты.

Чтобы устранить отраженную волну, вблизи нагрузки помещают *согласующее устройство*, параметры которого подбирают таким образом, чтобы оно трансформировало сопротивление нагрузки в волновое сопротивление, т.е. чтобы входное сопротивление устройства было равно волновому. В этом случае на линии до согласующего устройства будет существовать режим чисто бегущих волн.

Согласующее устройство не должно вносить потерь, следовательно, оно должно представлять собой реактивный четырехполосник. В качестве согласующих устройств чаще всего используются четвертьволновые трансформаторы и реактивные шунты.

На практике режим распространения электромагнитных волн в ЛП принято характеризовать коэффициентом бегущей волны $K_{БВ}$, который численно равен отношению минимальной амплитуды напряжения (или тока) в ЛП к максимальной.

Величина, обратная $K_{БВ}$, называется коэффициентом стоячей волны ($K_{СВ}$): $K_{СВ} = K_{БВ}^{-1}$. Между $K_{СВ}$ и модулем коэффициента отражения $|p|$ существует простая взаимосвязь:

$$K_{СВ} = \frac{1 + |p|}{1 - |p|}.$$

Для режима бегущих волн $|p| = 0$, $K_{\text{СВ}} = K_{\text{БВ}} = 1$. Для режима стоячих волн $|p| = 1$, $K_{\text{СВ}} = \infty$, $K_{\text{БВ}} = 0$. В режиме смешанных волн $0 < |p| < 1$, $1 < K_{\text{СВ}} < \infty$, $0 < K_{\text{БВ}} < 1$.

Таким образом, в дополнение к коэффициентам p и $K_{\text{БВ}}$ режим в линии вполне может быть охарактеризован значением $K_{\text{СВ}}$, который в зарубежной литературе именуется как VSWR (Voltage Standing Wave Ratio).

9.3.2 Двухпроводные и коаксиальные линии

Двухпроводные (симметричные) линии могут выполняться как в открытом, так и в экранированном варианте [36].

В современных системах связи открытые фидеры используются в основном для питания симметричных длинноволновых и средневолновых антенн. Открытый фидер либо крепится на жестких изоляторах, либо фиксируется при помощи изолированных оттяжек.

Коаксиальный (несимметричный) фидер может быть жестким или гибким. В жестком коаксиальном фидере внутренняя жила крепится на диэлектрических шайбах или на металлических изоляторах.

Гибкий коаксиальный фидер состоит из гибкой наружной металлической оплетки и внутреннего провода, поддерживаемого изоляторами либо погруженного в диэлектрик. Чаще всего изоляция выполняется в виде сплошного заполнения диэлектриком типа полиэтилена. Длина волны в кабеле с диэлектрическим заполнением рассчитывается по формуле $\lambda_k = \frac{\lambda}{\sqrt{\epsilon}}$, где λ — длина волны в свободном пространстве, ϵ — диэлектрическая постоянная. Соответственно, коэффициент уменьшения длины волны $\xi = \sqrt{\epsilon}$.

Таблица 9.1. Электрические характеристики гибких коаксиальных кабелей

Тип кабеля	Волновое сопротивление, Ом	Затухание при $f = 100$ МГц	Внешний диаметр, мм	Рабочее напряжение, кВ
РК-29	50	0,113	$9,8 \pm 0,5$	1,5
РК-19	52	0,2	$4,2 \pm 0,2$	1,0
РК-6	52	0,05	$12,4 \pm 0,6$	4,5
РК-47	52	0,08	$10,3 \pm 0,5$	1,0
РК-49	72	0,13	$6,9 \pm 0,3$	1,0
РК-3	75	0,07	$13,0 \pm 0,7$	5,5
РК-1	77	0,11	$7,3 \pm 0,4$	3,0
РК-2	92	0,09	$9,6 \pm 0,5$	4,5

Важнейшей характеристикой коаксиальных линий является их экранирующая способность, показывающая, во сколько раз мощность сигнала, распространяющегося во внутреннем проводе, превышает мощность той его части, которая просачивается за пределы экранирующей «рубашки». С другой стороны, эта же

характеристика показывает, во сколько раз мощность наведенной во внутреннем проводе помехи меньше помехи, действующей в свободном пространстве. Для лучших образцов гибких коаксиальных кабелей с двойным экраном ослабление внешних сигналов может достигать 90 дБ, а для жестких коаксиалов — до 110 дБ. В то же время простые кабели с одинарной экранирующей оплеткой подавляют внешнюю помеху лишь на 30–40 дБ.

9.3.3 Прямоугольные и круглые волноводы

Волновод — это полый металлический либо металлизированный трубопровод, используемый для передачи электромагнитных волн [36, 79]. В качестве волновода можно использовать трубу с любой формой поперечного сечения. Конструктивно более удобны прямоугольные волноводы. Круглые волноводы используются чаще всего там, где необходима осевая симметрия волновода, например во вращающихся сочленениях. Всякий волновод является своеобразным фильтром верхних частот, поскольку вдоль него могут распространяться только те колебания, чья длина волны меньше некоторого критического значения $\lambda_{кр}$.

Величина $\lambda_{кр}$ определяется поперечными размерами волновода и типом колебания, распространяющегося в нем. Чем меньше поперечные размеры и чем сложнее картина поля, тем меньше критическая длина волны. Обычно $\lambda_{кр}$ имеет тот же порядок, что и поперечный размер волновода, поэтому волноводы целесообразно применять для передачи волн дециметрового, сантиметрового и миллиметрового диапазонов.

В любом волноводе устанавливается такая картина поля, при которой на стенках волновода касательная составляющая электрического и нормальная составляющая магнитного полей равны нулю.

Вдоль волновода могут распространяться поперечно-электрические колебания с продольной составляющей вектора напряженности магнитного поля, которые называют волнами типа *TE* (transverse electric) или волнами *H*-типа, а также волны поперечно-магнитные с продольной составляющей вектора напряженности электрического поля. Последние называют волнами типа *TM* (или волнами *E*).

Волна, которая может распространяться в волноводе уединенно, без примеси других волн, называется основной для волновода, а все остальные именуются волнами высших типов. Обычно волновод должен работать на основной волне без примеси высших волн. С этой целью размеры волновода выбираются так, чтобы выполнялось неравенство $\lambda_{кр.осн} > \lambda_{кр} > \lambda_{кр.высш.}$

Если для выбранного типа колебаний соблюдается условие $\lambda < \lambda_{кр}$, то в волноводе, как и в любой фидерной линии, в зависимости от характера нагрузки могут существовать режимы бегущих, стоячих и смешанных волн.

Прямоугольный волновод

В прямоугольном волноводе (рис. 9.7, а) могут существовать волны *E* и *H*. Критическая длина волны для волн *E* и *H* определяется соотношением [79]:

$$\lambda_{кр} = \frac{2}{\sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2}},$$

где m, n — целочисленные положительные индексы, характеризующие тип волны и указывающие количество полуволн, укладываемых вдоль соответствующих сторон волновода (рис. 9.7, б).

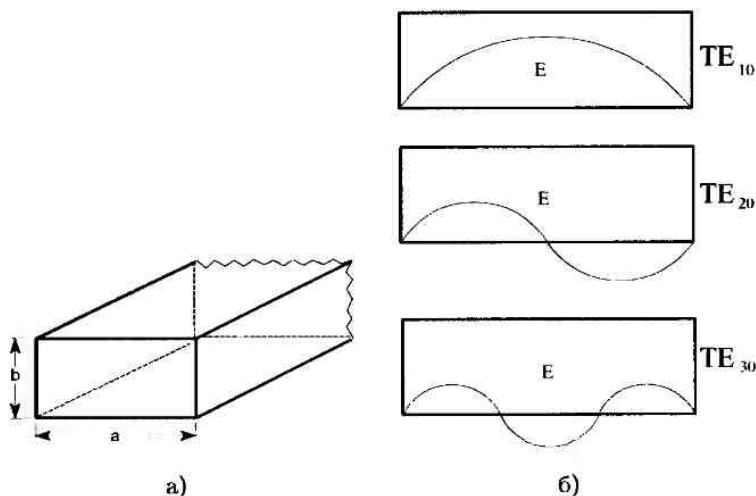


Рис. 9.7. Прямоугольный волновод (а) и примеры распространяющихся в нем волн в поперечном сечении (б)

Для волн E_{mn} : $m = 1, 2, 3, \dots$; $n = 1, 2, 3, \dots$

Для волн H_{mn} : $m = 0, 1, 2, 3, \dots$; $n = 0, 1, 2, 3, \dots$

Для волны H_{mn} значения $m = 0$ и $n = 0$ совместно невозможны. Наибольшее распространение получили прямоугольные волноводы с размерами стенок $a \sim 2b$. Для всех случаев, когда $a > b$, наибольшая величина $\lambda_{кр}$ соответствует волне H_{10} ($m = 1$; $n = 0$), которая поэтому является основной волной прямоугольного волновода. Для нее $\lambda_{кр} = 2a$. Основные данные прямоугольных волноводов приведены в табл. 9.2.

Таблица 9.2. Прямоугольные волноводы

Внутренние размеры, мм	Полоса пропускания, см	Толщина стенок, мм	Предельная мощность, МВт	Максимальное затухание, дБм
146 × 292	33–46,8	3,18	45,5	0,003
98 × 196	20,5–31,2	3,18	18,9	0,005
65 × 130	13,63–20,7	2,03	8,3	0,012
45 × 90	9,1–13,83	2,03	3,9	0,017
34 × 72	7,6–11,35	2,03	2,4	0,024
24 × 43	5,01–7,62	1,63	1,04	0,044
15 × 35	3,67–5,58	1,63	0,54	0,072
10 × 23	2,5–3,66	1,27	0,23	0,127

Круглые волноводы

В круглых волноводах также могут существовать E и H волны. Критическая длина волны [36]

$$\lambda_{кр} = \frac{2\pi a}{U},$$

где a — радиус волновода;

$$U = \begin{cases} U_{mn}, \\ U'_{mn}, \end{cases}$$

U_{mn} — для волн E ; U'_{mn} — для волн H ; U_{mn} — n -й корень функции Бесселя первого рода m -го порядка; U'_{mn} — n -й корень производной функции Бесселя первого рода m -го порядка.

Возможные значения индексов: $m = 0, 1, 2, 3, \dots$; $n = 1, 2, 3, \dots$. Индекс m указывает на периодичность поля волны (число «стоячих волн» по периметру окружности волновода). Случай $m = 0$ соответствует волне с осевой симметрией.

Индекс n указывает на число независимых коаксиальных областей, в каждой из которых энергия «циркулирует» как в обособленном волноводе, не затекая в соседнюю область.

Основной волной круглого волновода является волна H_{11} (рис. 9.8, а). Для нее $\lambda_{кр} = 3,41a$. Волна H_{11} используется в устройствах, где необходимо осуществлять поворот плоскости поляризации, а также в различных волноводных переходах.

Наряду с волной H_{11} применяется симметричная волна типа H_{11} (рис. 9.8, б). Она обладает малым затуханием, убывающим с ростом частоты, что позволяет использовать ее в волноводных линиях дальней передачи.

Симметричная E_{01} волна используется во вращающихся сочленениях. В них контакт наружных стенок достигается за счет четвертьволнового коаксиального отрезка, обеспечивающего электрический контакт в месте вращения.

Следует отметить, что волноводная техника находит применение преимущественно в радиорелейном оборудовании и в СВЧ-аппаратуре базовых станций. В малогабаритных абонентских терминалах она не применяется.

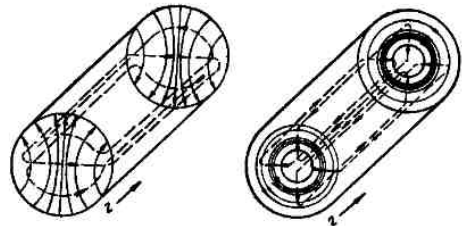


Рис. 9.8. Волны типа H_{11} (а) и H_{11} (б) в круглом волноводе

9.3.4 Направленные ответвители и волноводные мосты

Направленные ответвители — это элементы волноводных фидерных линий, предназначенные для передачи части электромагнитной энергии из одного волновода (основного) в другой (вспомогательный) [36]. В последнее время они широко используются в схемах линеаризации передаточных характеристик передающих каналов базовых станций сотовой связи. Если в основном волноводе существует волна одного направления, то во вспомогательном волноводе появится такая

же волна. В реальных ответвителях при этом дополнительно возникают небольшие по амплитуде электромагнитные колебания нежелательных направлений. Отношение мощности волн, распространяющихся во вспомогательном волноводе в требуемом и нежелательном направлениях, характеризует направленность ответвителя. В идеальном случае это отношение обычно составляет 20–40 дБ.

Связь между основным и вспомогательным волноводами характеризуют переходным ослаблением, величина которого может принимать значения от 0 дБ (полная передача энергии из основного волновода во вспомогательный) до 50–70 дБ, когда во вспомогательный волновод передается только часть основной мощности (10^{-5} – 10^{-7}).

Различают несколько разновидностей ответвителей. Ответвители с четверть-волновым разномом элементов связи в простейшем случае выполняются двухдырочными с отверстиями связи в узкой стенке прямоугольного волновода, работающего на основной волне. В них направленность достигается за счет того, что волны, прошедшие через отверстия, в нужном направлении складываются в фазе, а для обратного (нежелательного) направления за счет разности электрических путей оказываются противофазными и взаимно компенсируются.

Для уменьшения переходного ослабления и увеличения направленности в диапазоне частот применяют ответвители с несколькими разнесенными отверстиями. Такой ответвитель можно рассматривать как каскад из нескольких двухдырочных ответвителей. Иногда вместо круглых отверстий применяют четверть-волновые шлейфы, связывающие широкие стенки.

Ответвители с распределенной связью конструктивно выполняют в виде двух волноводов, соединенных через длинную щель в общей стенке, или в виде двух ленточных линий, расположенных достаточно близко одна к другой. Каждый участок щели (линии) можно рассматривать как элементарный возбудитель.

Волноводные мосты используются для развязки генераторов, имеющих разные частоты и работающих на общую нагрузку, для измерения рассогласований, а также в качестве элементов антенных коммутаторов и балансных смесителей. Наиболее часто применяются Т-мост и щелевой мост [36].

Т-мост (рис. 9.9, а) представляет комбинацию из двух волноводных тройников. Если к плечам 3 и 4 моста (рис. 9.9, а) подключить согласованные нагрузки и питать его со стороны плеча 2, то из-за симметрии моста СВЧ энергия в плечо 1 проникать не будет. При питании со стороны плеча 1 электромагнитная волна пойдет к плечам 3 и 4, не проникая в плечо 2. Таким образом, при симметричных нагрузках плеч 3 и 4 выходы 1 и 2 оказываются развязанными. Подача сигналов на входы 3, 4 позволяет снимать с плеча 2 разностный сигнал, а с выхода 1 — суммарный.

Наконец, если питать мост через плечо 2, то рассогласование в одном из плеч 3 или 4 вызывает появление СВЧ энергии в плече 1. Это явление позволяет использовать мост в качестве индикатора рассогласования. Недостатком Т-моста является необходимость применения дополнительного согласующего устройства и жесткие требования к точности изготовления.

Щелевой мост в наиболее распространенном виде имеет конструкцию из двух волноводов, имеющих общую узкую стенку, часть которой вырезана (рис. 9.9, б).

В центральной части стоит емкостный штырь, компенсирующий отраженные волны, возникающие в месте сочленения узкого и расширенного участков.

Щелевой мост можно рассматривать как направленный ответвитель с распределенной связью и переходным ослаблением 3 дБ. Если питать мост со стороны плеча 1 и в плечи 3 и 4 включить согласованные нагрузки, то в плечо 2 электромагнитная волна поступать не будет.

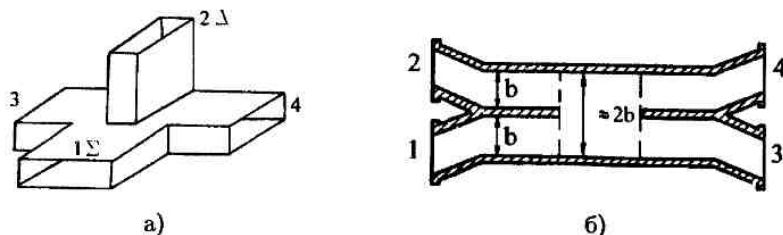


Рис. 9.9. Т-мост (а) и щелевой мост (б)

Длина щели $L_{\text{щ}}$ моста выбирается из условия [36]:

$$L_{\text{щ}} = \frac{1}{4} \frac{1}{\sqrt{1 - \left(\frac{\lambda}{4b}\right)^2} - \sqrt{1 - \left(\frac{\lambda}{2b}\right)^2}}.$$

В этом случае мощность, поступающая в плечо 1, делится поровну между плечами 3 и 4. При этом фазы колебаний в плечах будут отличаться на 90° . В плечо 2 сигнал не проходит. Если одна из нагрузок в плече 3 или 4 окажется несогласованной, то появляется отраженная волна, часть которой пойдет в плечо 2. Таким образом, по выполняемым функциям щелевой мост аналогичен Т-мосту, однако достоинством щелевого моста являются простота конструкции и лучшие диапазонные свойства.

9.3.5 Ферритовые фазовращатели и циркуляторы

В СВЧ линиях передачи получили широкое распространение устройства, в которых в качестве диэлектрика используются анизотропные материалы, обладающие различными свойствами в отношении волн, распространяющихся в разных направлениях. Наиболее часто в качестве подобного материала до сих пор использовался намагниченный феррит, хотя в последнее время, с открытием полимерных магнитов, этот подход пересматривается.

Феррит представляет собой магнитодиэлектрик с кристаллической структурой, напоминающий по внешнему виду керамику [36, 48, 79]. Анизотропия намагниченного феррита проявляется в том, что волны с различным направлением вращения плоскости поляризации распространяются в феррите с различными фазовыми скоростями и по-разному поглощаются. Изменяя подмагничивающее поле, можно менять соотношения параметров распространения для волн прямого и обратного направлений.

В линиях передачи, заполненных ферритом, наиболее часто используются следующие эффекты:

- не­взаимный (т. е. разный для прямых и обратных волн) поворот плоскости поляризации;
- не­взаимное фазовое запаздывание;
- не­взаимное поглощение.

Устройства, порождающие не­взаимный поворот плоскости поляризации, называются *не­взаимными вращателями*. Устройство, обеспечивающее не­взаимное фазовое запаздывание волн, именуют *не­взаимным фазовращателем*.

Наиболее удобны фазовращатели в виде отрезка прямоугольного волновода с поперечно-намагниченной ферритовой пластиной.

Разница фазовых запаздываний для волн противоположного направления определяется величиной подмагничивающего поля, пронизывающего феррит. При этом напряженность магнитного поля берется много меньше резонансной, чтобы поглощение было примерно одинаковым и малым для волн как прямого, так и обратного направления.

Устройства, по-разному поглощающие прямые и обратные волны, называются *не­взаимными ослабителями* (изоляторами отраженной волны).

Простейший ослабитель можно получить, если в рассмотренном выше устройстве в виде отрезка прямоугольного волновода с намагниченным ферритом увеличить значение подмагничивающего поля до величины, соответствующей условию ферромагнитного резонанса и интенсивного поглощения для одной из волн (например, отраженной). При этом прямая волна будет проходить с малым затуханием.

Последнее из рассматриваемых здесь СВЧ-устройств — циркулятор. Это волноводное устройство, в котором вследствие использования анизотропных элементов путь распространения электро­магнитной волны зависит от входа, на который она подается [36, 79]. Циркулятор с четырьмя питающими плечами схематически изображен на рис. 9.10.

Если питать его со стороны плеча 1, то сигнал поступает на выход 2. В то же время при питании циркулятора со стороны плеча 2 сигнал поступит уже не в плечо 1, а на выход 3 и т. д.

Циркулятор можно использовать в качестве антенного переключателя, если передатчик подключить к плечу 1, антенну — к выходу 2, приемник — к плечу 3, а к выходу 4 — поглощающую согласованную нагрузку.

Основными частями любого циркулятора являются щелевые мосты и фазовращатели. Наибольшую известность получили циркуляторы, использующие эффект не­взаимного фазового сдвига.

В заключение следует отметить, что рассмотренные варианты построения фидерных линий и их элементов не исчерпывают всего многообразия устройств подобного типа. За более подробными сведениями читатель может обратиться к специальной литературе, например [11, 36].

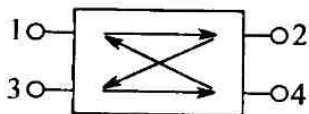


Рис. 9.10. Варианты прохождения сигнала в циркуляторе

9.4 Основные понятия теории и техники антенн. Классификация антенных систем

Антенные системы являются радиотехническими устройствами, предназначенными для излучения и приема электромагнитных волн с требуемой пространственно-частотной и поляризационной избирательностью.

По сути, в системах связи они служат конечным пунктом на пути формирования излучаемых радиосигналов и входным устройством при их приеме, в силу чего антенны достаточно сильно влияют на конечные характеристики систем связи.

С точки зрения физики антенны выполняют роль связующего звена между электромагнитными волнами, распространяющимися в среде, и электрическими токами, протекающими в линиях передачи (фидерах). В широкополосных системах передачи информации используется большой ассортимент антенн. В процессе своего развития антенны постоянно усложнялись, появились принципиально новые их конструкции, расширился спектр выполняемых ими функций. Чтобы систематизированно описать все их многообразие, удобно прибегнуть к классификации антенных систем, используя наиболее характерные их признаки.

Все антенны можно разбить на два основных класса: антенны, излучающие энергию в значительную часть пространства (слабонаправленные или «ненаправленные» антенны), и антенны, концентрирующие энергию в малом объеме (направленные антенны). Важным признаком в классификации антенн является также их геометрия. В соответствии с используемыми при проектировании геометрическими формами будем различать антенны евклидовой и неевклидовой геометрии. К последнему классу антенн относятся появившиеся сравнительно недавно так называемые фрактальные антенны.

Развитие микропроцессорной техники в последние десятилетия привело к тому, что многие операции обработки сигналов, в том числе адаптивной, все чаще реализуются непосредственно в антенной системе. С другой стороны, процесс придания направленных свойств антенне может быть также полностью либо частично реализован в цифровом виде, при этом появляется возможность адаптивно изменять характеристики излучения антенной системы под конкретные нужды пользователя. В результате такого процесса в антенной технике стали различать обычные и «интеллектуальные» антенны, получившие за рубежом наименование Smart-антенн.

Обычные антенные устройства, находящие применение в настоящее время, весьма разнообразны как с точки зрения выполняемых задач, так и с точки зрения их конструкции. Несмотря на это, положив в основу деления антенн конструктивное исполнение и принцип действия, их можно разделить на несколько основных типов [47].

Проволочные вибраторные антенны выполнены из тонких (по сравнению с длиной волны) или толстых проводников в виде труб, диаметр которых соизмерим с длиной волны.

Простейшими проволочными антеннами да и антеннами вообще можно считать элементарный вибратор и элементарную рамку. В самом распространенном

случае проволочная антенна представляет собой прямолинейный провод длиной около полуволны, разделенный на две равные части, к которым подключаются провода питающей линии. Такая антенна называется симметричным полуволновым вибратором (рис. 9.11).

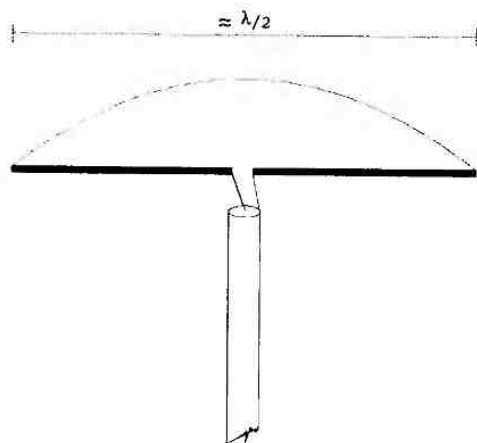


Рис. 9.11. Симметричный полуволновый вибратор

В диапазоне средних и длинных волн в качестве антенн используются несимметричные вертикальные вибраторы, большей частью питаемые у основания. Характеристики излучения данных антенн близки к свойствам элементарного вибратора (рис. 9.12). Для получения направленного излучения могут использоваться сложные антенны, состоящие из большого числа вибраторов. Примером такого рода антенн является антенна Уда-Яги (рис. 9.13).



Рис. 9.12. Характеристика излучения несимметричного вертикального вибратора

Щелевые (дифракционные) антенны представляют собой одно или несколько отверстий в виде узких щелей, прорезанных в плоскости антенны, в стенках волновода либо объемного резонатора. Щели формируют таким образом, чтобы они пересекали линии поверхностных токов, возбуждаемых на несущей поверхности. Тем самым нарушается экранировка, и возникает излучение электромагнитных волн в окружающее пространство. Чтобы получить высоконаправленное излучение, используются системы щелей.

Антенны акустического типа. К антеннам данной группы относятся рупорные антенны, применяемые главным образом в сантиметровом диапазоне волн. Чтобы получить рупорную антенну, достаточно, например, плавно расширить все четыре или одну из пар противоположных стенок волновода прямоугольного сечения. В первом случае получается пирамидальный рупор (рис. 9.14), во втором — секториальный. При расширении волновода круглого сечения формируется конический рупор.

Антенны оптического типа. К антеннам этого класса относятся зеркальные и линзовые антенны, аналоги которых можно найти в оптике. В антеннах оптического типа сравнительно широкая диаграмма направленности первичного источника электромагнитных волн, так называемого облучателя, преобразуется в узкую диаграмму направленности той или иной формы за счет отражения излучения от зеркала или преломления в линзе.

Антенны поверхностных волн. Основой для создания таких антенн служат направляющие системы, вдоль которых могут распространяться поверхностные электромагнитные волны. В качестве направляющих систем могут быть использованы диэлектрические стержни круглого сечения, металлические стержни, покрытые слоем диэлектрика, металлические ребристые структуры и т. п. Характерной особенностью антенн поверхностных волн является излучение электромагнитных волн преимущественно вдоль оси. По этой причине данные антенны называются антеннами осевого излучения.

В зависимости от ориентации вектора напряженности электрической составляющей электромагнитного поля, определяющей так называемую *поляризацию* излучения, при классификации антенн различают антенны с линейной и вращающейся (круговой либо эллиптической) поляризацией. Линейная поляризация, обычно вертикальная, используется, например, в стандарте IEEE 802.11b (в зависимости от положения вибраторного излучателя, рис. 9.15). Ниже, на рис. 9.16, показан пример электромагнитной волны с круговой поляризацией, у которой вектор напряженности электрического поля E совершает вращательные движения. Очевидно, что максимум мощности принятого сигнала достигается, когда участвующие в процессе передачи и приема информации антенны используют идентичную поляризацию. Поэтому в системе связи поляризация излучений приемной и передающей антенн должна быть строго согласована.

Антенны с вращающейся поляризацией поля в простейшем случае можно получить из двух полуволновых вибраторов, если их оси расположить взаимно перпендикулярно, а токи запитки сдвинуть по фазе на 90° . Подобного же типа антенну можно выполнить в виде рупорного излучателя. Специальным типом антенны с круговой поляризацией поля является *спиральная антенна*, представляющая собой фрагмент спирали, расположенный в одной плоскости либо навитый вокруг заданной оси.

В большинстве известных схем классификации особую группу антенн составляют антенны с управляемым направлением интенсивности излучения. В общем случае такие антенны представляют собой системы излучателей, расположен-

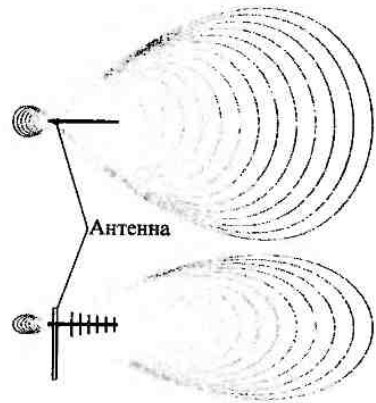


Рис. 9.13. Антенна Уда-Яги

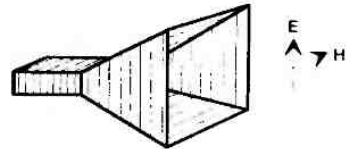


Рис. 9.14. Пирамидальный рупор

ных в пространстве по какому-либо закону, их называют *антенными решетками*. Простейшим примером таковых являются *фазированные антенные решетки (ФАР)*. Их отличительной особенностью является наличие общего для всех излучателей фазового центра — точки, относительно которой осуществляется фазирование токов запитки излучателей. В отличие от трактовки, приведенной в [92], следует указать, что фазовый центр может находиться в любой точке апертуры антенной решетки, а не только в ее середине. Здесь под апертурой антенны понимается любая геометрическая поверхность, через которую проходит вся излучаемая мощность [92].

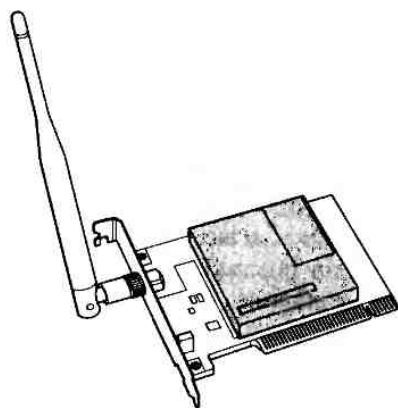


Рис. 9.15. PCI-карта приемопередатчика стандарта IEEE 802.11b

Далее подробно будут рассмотрены лишь те из упомянутых антенн, которые представляют наибольший интерес для решения задач современной широкополосной связи. Однако, прежде чем приступить к их описанию, имеет смысл остановиться на изложении основных понятий антенной теории, позволяющих объективно сопоставлять и оценивать качества различных антенных конструкций.

Поскольку работа любой антенны всегда связана с двумя видами энергии: токов высокой частоты и электромагнитного излучения, то в соответствии с этим ее параметры можно условно разделить на две группы. К первой из них относятся параметры, влияющие на энергию токов высокой частоты, а именно: входное сопротивление, резонансные частоты, частотная характеристика, полоса пропускания и т. п. Вторую группу составляют параметры, определяемые наличием энергии электромагнитного излучения. Они образуют группу так называемых характеристик излучения, к которым относятся: диаграмма направленности, ширина ее главного лепестка, относительный уровень побочных максимумов, коэффициент усиления, коэффициент направленного действия, эффективная площадь и др. [97].

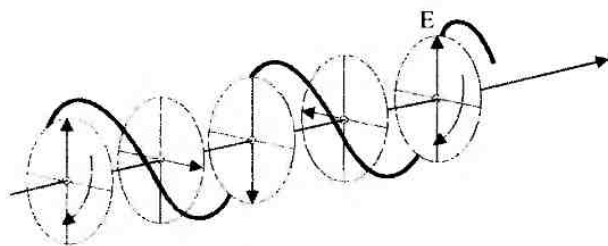


Рис. 9.16. Пример волны с круговой поляризацией (E — вектор напряженности электрического поля)

Обе группы параметров характеризуют работу в равной мере как передающих, так и приемных антенн. Согласно теореме взаимности и ее следствиям пара-

метры антенны имеют определенную величину независимо от того, используется ли антенна в качестве передающей или приемной.

На практике определить некоторые параметры антенн можно лишь путем непосредственных измерений. Такие параметры называют первичными. Остальные же, именуемые вторичными, могут быть найдены графически или расчетным путем по известным первичным. В первой из указанных выше групп параметров первичным является входное сопротивление. По его известным значениям в диапазоне частот можно найти полосу пропускания, резонансные частоты и частотную характеристику. Во второй группе первичными параметрами являются диаграмма направленности и коэффициент усиления. Зная диаграмму направленности, можно определить ширину главного лепестка, уровень боковых максимумов и получить представление о коэффициенте направленного действия. Зная коэффициент усиления, можно вычислить эффективную площадь.

В большинстве случаев антенна представляет собой четырехполюсник, одна пара полюсов которого является эквивалентной моделью свободного пространства. Как и любой четырехполюсник, антенна характеризуется входным сопротивлением.

Входное сопротивление антенны — это отношение напряжения источника электродвижущей силы (э.д.с.), включенного на входе антенны, ко входному току. Оно обозначается через Z_A . В общем случае входное сопротивление является комплексной величиной, его активная и реактивная составляющие обозначаются через R_A и X_A . Если $X_A < 0$, то реактивная составляющая является емкостной, в противном случае ($X_A > 0$) — индуктивной.

Пространственное распределение напряженности излучаемого антенной электромагнитного поля можно описать некоторой функцией, которая называется **диаграммой направленности (ДН)** и представляется обычно в виде одного или нескольких сечений в главных плоскостях. Если говорить более точно, диаграмма направленности антенны в определенной плоскости — это функция зависимости амплитуды напряженности поля от пространственной координаты, в качестве которой, как правило, рассматривается угол отклонения от нормали к плоскости раскрытия антенны. В общем случае ДН является комплексной величиной. Зачастую при построении диаграмм вместо напряженности поля рассматривают излучаемую мощность.

Строго говоря, прежде в теории антенн различали также понятие характеристики направленности (направленного действия, ХНД), представлявшей собой формульную зависимость величины энергии от направления излучения. При этом графическое изображение ХНД называлось уже упомянутой диаграммой направленности антенны. Однако со временем эти понятия стали отождествляться, и в последнем справочнике по антенной технике [92] приведено лишь понятие диаграммы направленности.

ДН могут изображаться в трехмерном (рис. 9.17) и двумерном виде. При этом используются полярные и прямоугольные системы координат в линейном либо логарифмическом масштабах.

Линейные антенны, подобно всем линейным цепям, удовлетворяют принципу взаимности. Как входное сопротивление (со стороны линии передачи), так и диаграмма направленности линейной антенны не зависят от того, работает ли данная

антенна в режиме передачи или приема. Большинство антенн может быть отнесено к линейным, но некоторые из них могут содержать материалы с нелинейными характеристиками, такие как полупроводники и ферриты, аналого-цифровые преобразователи. В этих случаях надо специально оценивать возможность применения принципа взаимности. Например, в цифровых антенных решетках, о которых пойдет речь ниже, при так называемом цифровом формировании диаграммы направленности и сверхрешеточном разрешении сигналов по направлению прихода широко используются нелинейные математические операции. Как следствие, результирующие ДН подобных приемных антенн будут крайне узконаправленными и остроконечными. При передаче воспроизвести данные характеристики направленности не удастся. В этом случае принцип взаимности нарушается.

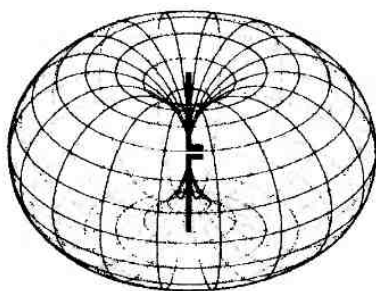


Рис. 9.17. Пример трехмерной ДН

Для описания свойств направленных антенн по их диаграммам направленности существует несколько параметров. Это прежде всего ширина луча (ширина главного лепестка диаграммы направленности) по половинной мощности, т. е. угловое расстояние между точками, в которых мощность излучения составляет половину от максимальной. Участки диаграммы направленности, не относящиеся к главному лепестку, называются боковыми лепестками (рис. 9.18).

Важными параметрами на практике служат уровни боковых лепестков и форма их огибающей. Уровень боковых лепестков равен отношению пиковой величины наибольшего из них (обычно соседнего с главным лепестком) к максимуму главного лепестка и выражается, как правило, в децибелах.

Форма ДН антенн реально сильно зависит от положения окружающих предметов в пространстве и электрических характеристик подстилающей поверхности, поэтому, как правило, понятие ДН является весьма условным и идеализированным.



Рис. 9.18. Структурные элементы типовой диаграммы направленности

Ширина главного луча и уровни боковых лепестков определяются геометрической формой и размерами антенны. Так, антенны с широким лучом соизмеримы по габаритам с рабочей длиной волны, в то время как антенны с узким лучом многократно ее превышают.

Способность антенны концентрировать излучаемую энергию в пространстве или, наоборот, наиболее эффективно принимать энергию, приходящую из некоторой части пространства, характеризуется **коэффициентом направленного действия (КНД)** [81]. Он определяется как отношение мощности, приходящейся на единицу телесного угла в направлении максимального излучения, к средней мощности излучения на единицу телесного угла. Если функция $F(\theta, \beta)$ — двумерная ДН антенны, а направление максимального излучения определяется углами θ_0, β_0 , то коэффициент направленного действия G будет равен [81]:

$$G = \frac{4\pi[F(\theta_0, \beta_0)]^2}{P},$$

где P — полная излучаемая антенной мощность.

КНД антенны обычно оценивают в децибелах по отношению к КНД изотропного излучателя (гипотетической антенны, излучающей равномерно во всех направлениях и нереализуемой физически).

По определению, КНД изотропного излучателя равен 1, что составляет 0 дБ. Иногда КНД антенны оценивают путем сравнения с коэффициентом направленного действия элементарного излучателя, равным 1,5 или 1,76 дБ, либо с КНД полуволнового вибратора (рис. 9.11), равным 1,64 (2,14) дБ.

Используя понятие об изотропном излучателе, можно дать иное определение КНД — это отношение максимальной интенсивности излучения данной антенны к интенсивности излучения изотропной антенны при условии, что обе антенны излучают одинаковую полную мощность. В данном случае под интенсивностью излучения следует понимать мощность, приходящуюся на единицу телесного угла. Таким образом, КНД показывает, во сколько раз возрастает мощность, излучаемая в выбранном направлении, по сравнению с мощностью ненаправленного излучения.

Как следует из определения, КНД не учитывает потерь в проводниках и диэлектриках, потерь за счет рассогласования питающих линий с нагрузкой. Для учета этих и других потерь был введен другой параметр — коэффициент усиления антенны. **Коэффициент усиления** равен отношению мощности единицы телесного угла, излучаемой в направлении максимума электромагнитного поля, к средней мощности единицы телесного угла, подводимой к зажимам антенны. Коэффициент усиления обычно используется для оценки качества передающих антенн, однако упомянутый выше принцип взаимности позволяет рассматривать данный показатель и в отношении приемных антенн. В этом случае в приведенном определении следует говорить об отношении мощности на единицу телесного угла, принимаемой с направления максимума поля, к средней мощности на единицу телесного угла, снимаемой с выходных зажимов антенны.

Непосредственной характеристикой приемных антенн является их **эффективная площадь**, под которой понимают площадь эквивалентной плоской антенны, поглощающей из приходящей электромагнитной волны с плоским волновым фронтом столько же энергии, сколько ее поглощает и данная антенна. Между эффективной площадью A_e и КНД антенны G существует однозначная связь: $G = 4\pi A_e / \lambda^2$, где λ — длина волны излучения [81].

Практически может оказаться, что антенна будет иметь поляризацию, отличающуюся от поляризации падающей волны, или не будет согласована с нагрузкой. В этих случаях полезная поглощаемая мощность будет уменьшаться, а эффективная площадь антенны — падать.

Приведенные соотношения играют важную роль в технике связи. Например, с их помощью можно выразить отношение мощностей излученного и принятого антеннами сигналов при передаче на большое расстояние R [81]:

$$\frac{P_R}{P_T} = \frac{A_R A_T}{\lambda^2 R^2},$$

где P_R — принимаемая мощность; P_T — излучаемая мощность; A_R , A_T — эффективные площади приемной и передающей антенн.

Данная формула справедлива при условии, что расстояние R между антеннами позволяет принять допущение о нахождении каждой из антенн в дальней зоне другой антенны. Это условие почти всегда выполняется в практике дальней связи. Если же расстояние окажется меньше (например, в случае двух ноутбуков, расположенных на удалении 1–2 м и обменивающихся информацией по протоколу 802.11b), необходимо пользоваться более сложными зависимостями.

9.5 Антенные системы евклидовой геометрии

Рассматривать конструктивные особенности различных типов антенн и их характеристик имеет смысл, руководствуясь принципом «от простого к сложному». Поэтому в ряду антенн евклидовой геометрии начнем с элементарных антенн.

9.5.1 Элементарные антенны (проволочные вибратор и рамка)

Элементарным вибратором или вибратором Герца является отрезок тонкого проводника, по которому течет переменный ток с постоянными амплитудами и фазой и длина которого много меньше, чем длина волны излучения (рис. 9.19, а).

Вектор напряженности электрического поля вибратора лежит в плоскости, проходящей через ось вибратора и точку наблюдения; вектор напряженности магнитного поля перпендикулярен этой плоскости. Ток в проводнике сопровождается излучением электромагнитных волн. Поле излучения элементарного вибратора можно разбить на три зоны: ближнюю, дальнюю и промежуточную. С точки зрения излучения наибольший интерес представляет дальняя зона. В ней электрическое и магнитное поля изменяются синфазно во времени, причем поверхностями одинаковых фаз являются сферы, общий центр которых находится в точке расположения вибратора. Амплитуды векторов напряженностей электрического и магнитного полей убывают обратно пропорционально первой степени расстояния.

Если вибратор находится в свободном пространстве, то мощность излучения такой антенны описывается формулой [47]:

$$P = 40\pi^2 I^2 \left(\frac{l}{\lambda}\right)^2,$$

где l — длина проводника, I — амплитуда тока в проводнике.

Существует понятие сопротивления излучения элементарного вибратора, под которым понимают такое активное (омическое) сопротивление R , на котором полностью поглощается мощность, равная мощности излучения вибратора [47]:

$$R = 80\pi^2 \left(\frac{l}{\lambda}\right)^2.$$

Произведение $\pi \frac{l}{\lambda}$ называется электрической длиной вибратора; следовательно, сопротивление излучения вибратора пропорционально квадрату его электрической длины.

Следует отметить, что рассмотренный элементарный вибратор является скорее теоретической абстракцией и служит базовым элементом для построения вибраторных антенн.

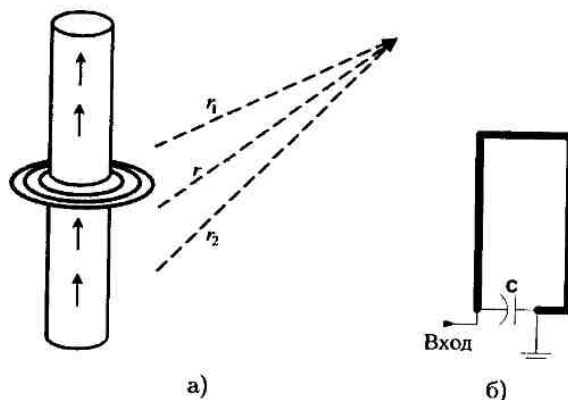


Рис. 9.19. Элементарные излучатели: вибратор (а) и рамка (б)

Другим элементарным источником электромагнитных волн в теории антенн является *элементарная рамка* [47], под которой понимают виток провода той или иной формы, проводящий переменный ток. Вариант ее подключения к источнику переменного тока приведен на рис. 9.19, б. Рамку можно считать элементарной, если длина ее контура значительно меньше длины волны, соответствующей частоте переменного тока. Сделанное допущение позволяет рассматривать ток в рамке как квазистационарный, т. е. считать, что вдоль провода рамки в каждый момент времени ток имеет одно и то же значение.

Сравнение рамки с элементарным вибратором показывает, что она существенно проигрывает ему по мощности излучения. Так, если мощность излучения вибратора длиной l пропорциональна $(l/\lambda)^2$, то для квадратной рамки со стороной l соответствующий коэффициент пропорциональности равен $(l/\lambda)^4$. По этой

причине рамки в качестве излучающих устройств, т.е. передающих антенн, как правило, не применяются. В основном они находят применение в роли приемных антенн при решении задач радиомониторинга, где направленные свойства рамок, не зависящие от их размеров, используются для целенаправленной передачи радиостанций. При этом для повышения эффективности рамки число ее витков увеличивают. Однако значительное увеличение количества витков оказывается невыгодным, поскольку приводит к росту реактивного сопротивления рамки, увеличению потерь в ее проводе и в диэлектрике (изоляции проводов, каркасе рамки и т. п.). Увеличение потерь с ростом частоты в большей степени обнаруживается в многovitковых рамочных антеннах.

9.5.2 «Ненаправленные» антенны

Под «ненаправленной» антенной, при всей условности этого термина, следует понимать антенну, ДН которой хотя бы в одной из угловых плоскостей не зависит от направления прихода (излучения) сигналов. Простейшей ненаправленной антенной является симметричный вибратор, состоящий из двух равных частей, присоединенных к клеммам генератора или линии передачи (рис. 9.11). Ненаправленные свойства ДН диполя проявляются в плоскости, перпендикулярной его оси. В плоскости же, проходящей через ось диполя, форма сечений ДН существенно зависит от соотношения его размеров и длины волны, иллюстрацией чему может служить рис. 9.20, где приведены частотно-зависимые ДН симметричного вибратора.

Активная часть входного сопротивления полуволнового симметричного вибратора равна 73,1 Ом, КНД составляет 1,64. Реактивная компонента входного сопротивления имеет индуктивный характер и равна 42,5 Ом. Следовательно, длина этого вибратора больше резонансной, поэтому для настройки в резонанс необходимо плечо вибратора укорачивать.

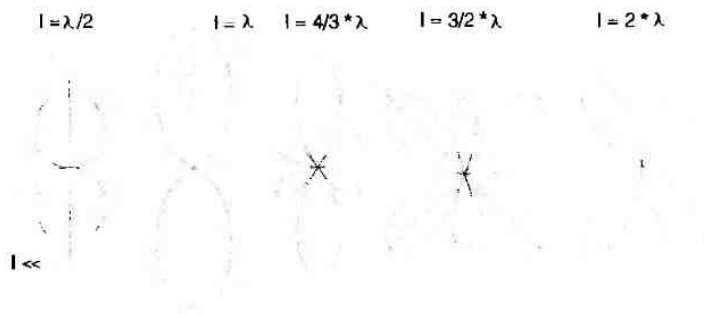


Рис. 9.20. Диаграммы направленности различных по длине симметричных вибраторов в плоскости их оси

На практике вместо симметричного вибратора зачастую используют несимметричный, именуемый в отличие от диполя монополю. Подобно диполю он также имеет ненаправленную ДН в плоскости, перпендикулярной монополю.

9.5.3 Направленные антенны

Простейший способ получения направленной антенны с пространственной избирательностью одновременно в азимутальной и угломестной плоскостях состоит в использовании системы из нескольких ненаправленных антенных элементов, расположенных и запитываемых по определенному закону. Примером такого рода является антенна типа «волновой канал», которую по именам предложивших ее более 80 лет назад японских создателей чаще называют антенной Уда-Яги. Она состоит из ряда параллельных вибраторов, расположенных перпендикулярно антенной оси (рис. 9.13). Если бы все вибраторы были активными, то для получения максимального излучения вдоль оси антенной системы и минимального излучения в противоположном направлении достаточно было бы расположить вибраторы на расстоянии четверти длины волны друг от друга. При этом их следует возбуждать так, чтобы ток в каждом последующем вибраторе (в сторону максимального излучения) отставал по фазе от тока в предыдущем антенном элементе на 90° при равенстве токов по амплитуде. Такая запитка (с точки зрения фазовых соотношений) может быть получена, если подключить вибраторы к двухпроводной линии через $0,25\lambda$. В антенне типа «волновой канал» активным является лишь один из вибраторов, а остальные — пассивные. Те из них, что расположены вдоль направления излучения, именуются директорами, а с противоположной от активного вибратора стороны — рефлекторами. Расстояние между активным вибратором и первым директором, как и расстояния между соседними директорами, выбирается примерно $0,1\lambda$ – $0,2\lambda$. Для антенны Уда-Яги характерен одиночный рефлектор, который устанавливается на расстоянии около $0,15\lambda$ – $0,2\lambda$ от активного вибратора (рис. 9.21) [47]. Длина рефлектора составляет около $0,53\lambda$. Директоры предназначены для формирования узкой диаграммы направленности, от них происходит еще одно название рассматриваемого типа антенн — директорные. Ток в директорах должен отставать по фазе от тока в активном вибраторе, по этой причине они делаются несколько короче — $0,5\lambda$. Число директоров может колебаться от 3 до 10, дальнейшее увеличение их количества не приводит к уменьшению ширины диаграммы направленности уже 15 – 20° [36, 47, 79].

Расчет антенны типа «волновой канал» сводится прежде всего к определению геометрических размеров антенны и токов в вибраторах. Такой расчет проводится для нескольких вариантов с выбором предпочтительного из них по критерию максимума КНД. Если же КНД антенны задан, то общую длину антенны, а по ней и число необходимых вибраторов можно найти из приближенного соотношения $G \approx 8L/\lambda$, где L — длина антенны [47].

Взаимные сопротивления Z_{mn} вибраторов определяются расстояниями между ними и их длиной. При числе директоров не больше трех расстояния между вибраторами следует все так же выбирать от $0,1\lambda$ до $0,2\lambda$, а длины вибраторов задавать такими, чтобы реактивное сопротивление рефлектора лежало в пределах от 0 до 40 Ом, а реактивное сопротивление директоров — в пределах от 0 до –40 Ом.

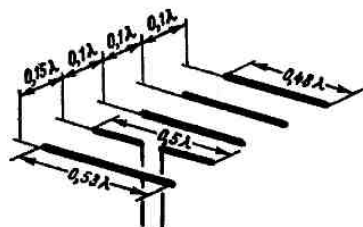


Рис. 9.21. Конструкция антенны Уда-Яги

При расчете диаграммы направленности антенна типа «волновой канал» условно рассматривается как антенна с бегущей волной тока. Такой подход связан с тем, что приближенно сдвиг фаз между токами в соседних вибраторах можно считать постоянным (в сторону запаздывания), а запитку антенны рассматривать как возбуждение волной, скорость распространения которой вдоль оси антенны несколько меньше скорости распространения в окружающем пространстве (сдвигу фаз в 90° в свободном пространстве соответствует расстояние $0,25\lambda$, а здесь это расстояние меньше — около $0,2\lambda$). Именно поэтому антенна названа «волновой канал».

Наряду со сложностью расчета подобного типа антенны обладают некоторыми недостатками. Один из них связан с малой диапазонностью, которая проявляется в резком росте максимумов боковых лепестков ДН и уменьшении КНД даже при небольшом изменении частоты несущей. С увеличением количества вибраторов диапазонность антенны ухудшается.

Еще один недостаток антенны Уда-Яги обусловлен тем, что благодаря влиянию рефлектора и директоров входное сопротивление активного вибратора может уменьшаться почти до 10 Ом, что затрудняет согласование его с питающим фидером, имеющим, как правило, волновое сопротивление в пределах от 50 до 75 Ом. По этой причине вместо обычного вибратора в антеннах типа «волновой канал» часто используются так называемые бивибраторы или шлейф-вибраторы, предложенные А.А. Пистолькорсом в 1936 году. Такой антенный элемент состоит из симметричного диполя (несколько короче полуволны), к которому в плоскости, перпендикулярной оси излучения, подключен второй вибратор такой же длины. Зачастую данный элемент именуют петлевым вибратором. Расстояние между его параллельными составляющими берется весьма небольшим — около $0,025\lambda$. Концы вибраторов соединяются перемычками, а питающий фидер подключается к клеммам диполя.

Другим вариантом директорной антенны, состоящей из вибраторных элементов, является логопериодическая антенна. Она отличается тем, что длины директоров по мере удаления от рефлектора убывают по логарифмическому закону.

Открытые на конце волноводы прямоугольного или круглого сечения могут рассматриваться как один из наиболее простых типов антенн с плоским излучающим раскрытием. Действие таких антенн сводится к тому, что поле волны, набегающей к открытому концу волновода, частично излучается в окружающее пространство. Небольшие размеры поперечного сечения волновода (по сравнению с длиной волны) приводят к тому, что ДН таких антенн оказывается сравнительно широкой. Данное обстоятельство ограничивает использование волноводов в качестве самостоятельных антенн, но в то же время они довольно часто применяются в сложных антеннах в качестве облучателей.

Недостатки волноводных излучателей, связанные с широкой ДН и наличием отраженной волны, могут быть в значительной степени устранены, если часть волновода, где происходит излучение, сделать расширяющейся. В результате этого получается так называемый пирамидальный рупор (рис. 9.14), который обладает лучшими направленными свойствами и выполняет роль согласующего устройства при переходе от обычного волновода к окружающему пространству.

Наряду с рупорами пирамидальной формы находят применение также секторальные рупоры. Если волновод прямоугольного сечения расширяется в плоско-

сти вектора напряженности электрического поля (здесь и далее считается, что в волноводе прямоугольного сечения распространяется основная волна типа H_{10}), то рупор называется *E*-секториальным. Если расширение волновода производится в плоскости вектора напряженности магнитного поля, то рупор будет *H*-секториальным. Как правило, волновод расширяют симметрично. Применяются, хотя и менее часто, рупоры круглого сечения, которые именуются коническими.

Расчет ДН рупорных антенн производится по распределению поля в раскрыве рупора и подробно рассмотрен в [47].

К достоинствам рупорных антенн следует отнести их хорошие диапазонные свойства, которые, по существу, определяются диапазонными свойствами волновода, запитывающего рупорную антенну. На практике достигается почти двукратное перекрытие по длине волны.

Кроме самостоятельного применения, рупорные антенны широко используются в качестве облучателей зеркальных, линзовых и других сложных антенн, а также в составе антенных решеток с управляемой ДН.

Зеркальными называются антенны, у которых поле в раскрыве формируется в результате отражения электромагнитной волны от металлической поверхности специального профиля. Кроме зеркала, в состав такой антенны входит облучатель.

Историю развития зеркальных антенн можно разделить на три периода. Первый из них связан с именем Герца, который в своих классических опытах по СВЧ-оптике в 1888 году применил параболический цилиндр в качестве фокусирующего зеркала для волны длиной 66 см [81]. Облучателем служил диполь, расположенный на фокальной линии. Апертура данной антенны имела размеры $1,2 \times 2$ м, что порождало сравнительно широкий луч. Вызывает удивление, что Герцу при таком широком луче удалось провести свои фундаментальные исследования.

Вслед за Герцем предпочтение зеркальной антенне в виде параболического цилиндра отдавали и другие исследователи, повторявшие или совершенствовавшие его опыты. В их числе был и Маркони, который для своих первых радиотелеграфных передач на волне длиной 25 см применял параболические цилиндры. К 1900 году переход радиотелеграфии на длинные волны, обеспечивающие связь на большие расстояния, ослабил интерес к сверхвысоким частотам и параболическим зеркалам [81].

Несмотря на перспективы, открывшиеся еще при зарождении радиотехники, далее последовал период застоя в развитии зеркальных антенн, продолжавшийся приблизительно до 1930 года. Лишь в 1931 году, когда была построена линия телефонной СВЧ-связи через Ла-Манш, интерес к параболическим антеннам вновь оживился. Данное событие можно считать началом современного периода в эволюции зеркальных антенн, продолжающегося и поныне.

Зеркала, применяемые в антенных системах СВЧ, имеют две основные геометрические характеристики: кривизну поверхности и форму граничной кривой. СВЧ-зеркала могут быть поляризованными, неполяризованными или поляризующими. В последнем случае зеркало является преобразователем вида поляризации волны.

Тип антенны определяется геометрической формой зеркала. В простейшем случае зеркало может представлять собой плоскую металлическую пластину достаточно больших размеров. Такая пластина играет роль рефлектора, благодаря

которому излучение будет происходить преимущественно по направлению нормали к поверхности зеркала. Плоское зеркало широко применяется как отдельно, так и в комбинации с другими зеркалами. Небольшие по размерам плоские зеркала используются в облучателях, а сравнительно габаритные — в качестве пассивных отражателей в радиорелейных системах, в том числе миллиметрового диапазона.

Несколько более сложной конструкцией является зеркало в виде двух плоских металлических пластин, чаще всего образующих прямой двугранный угол. Вместе с облучателем, представляющим собой симметричный вибратор, такое зеркало образует так называемую уголковую антенну. Вибратор обычно устанавливается в плоскости биссектрисы двугранного угла, образованного пластинами зеркала, параллельно его ребру.

Значительно более выраженной направленностью обладают антенны с зеркалом в виде параболоида вращения. Они имеют узкую ДН в двух плоскостях, которая называется ДН игольчатого типа. Величина отношения фокусного расстояния F к диаметру зеркала D у параболоидов обычно лежит в пределах от 0,2 до 1. Если нужно иметь антенну, ДН которой достаточно узкая в одной плоскости и широкая — в ортогональном сечении, то в качестве зеркала можно использовать усеченный параболоид вращения. Однако в такой антенне трудно получить ДН с большой разницей в ширине сечений в главных плоскостях. Поэтому для реализации «веерной» ДН чаще используется зеркало в виде параболического цилиндра с линейным облучателем.

Если параболический цилиндр ограничить двумя проводящими параллельными плоскостями, расположенными перпендикулярно фокальной линии, то получится сегментно-параболическая антенна типа «пилбокс» [81]. Термин «пилбокс» используется в случае, когда поляризация перпендикулярна ограничивающим плоскостям. Параболический цилиндр можно использовать в качестве облучателя другого параболического цилиндра, в результате чего образуется система скрещенных цилиндров [81]. Главное достоинство параболических цилиндрических систем состоит в простоте изготовления зеркала. Дополнительное преимущество связано с тем, что в системах такого типа сравнительно мала величина астигматизма.

Сферическое зеркало дает возможность смещать диаграмму направленности при движении облучателя по концентрической дуге сканирования [81]. При малых отношениях F/D из-за aberrаций по главной оси приходится использовать облучатели с коррекцией. В качестве другой граничной кривой может рассматриваться парабола. Если ее вращать относительно оси, расположенной позади вершины, то получится вазоподобный отражатель, по форме немного напоминающий песочные часы. Окруженный кольцом облучателей в плоскости наименьшего поперечного сечения, такой отражатель позволяет реализовать всенаправленное излучение в азимутальной плоскости.

Окружность и парабола являются наиболее характерными кривыми для конструкций зеркальных СВЧ-антенн в системах связи, поскольку для отражателей конические сечения имеют основное значение. Кроме параболических зеркал, находят применение эллиптические и эллипсоидные, а также гиперболические и гиперболоидные отражатели.

По количеству используемых зеркал все антенны данного класса, применяемые в системах широкополосной передачи, можно условно разделить на одно- и двухзеркальные.

Однозеркальными системами называются такие антенны, в которых для получения требуемой диаграммы направленности используется только одна отражающая поверхность (рис. 9.22). При этом антенна может иметь апертуру любой формы, простую или сложную поверхность, работать с одним или несколькими облучателями. Для однозеркальной системы важно лишь то, чтобы формирование луча происходило только после одного отражения. Этот тип зеркальных систем наиболее распространен на практике и применяется для создания самых различных диаграмм направленности.

Например, в однозеркальной системе можно создать веерный луч, для чего годится любой из четырех методов: 1) ограничение раскрыва в одном направлении; 2) использование зеркала, у которого фокусировка осуществляется только в одной плоскости; 3) применение системы распределенных облучателей и 4) любая комбинация трех указанных методов. Выбор того или иного варианта зависит от множества факторов, среди которых обычно наиболее важно соотношение размеров луча в двух главных плоскостях.

Если это отношение должно быть большим, то простейшим способом его достижения является ограничение размеров раскрыва в одном направлении и использование зеркала типа «пилбокс». В такой системе можно получить отношение диаметров сечений веерной диаграммы направленности более $100 : 1$. Для создания веерного луча с меньшей неоднородностью сечений можно применить антенну с параболическим цилиндром в качестве отражателя. Эта конструкция является примером зеркальной антенны, в которой идеальное фокусирование луча происходит только в одной плоскости.

К **двухзеркальным** относятся антенны, в которых для получения направленного излучения используются две отражающие поверхности. В обычном случае энергия входящей волны отражается от большого зеркала (первичный отражатель) на меньшее зеркало (вторичный отражатель) и затем попадает на облучатель. Удобство двухзеркальной конструкции состоит в том, что облучатель и связанные с ним электронные блоки размещаются сзади большого зеркала, как правило, в его вершине, хотя облучатель может также находиться между обоими зеркалами.

Наиболее распространенной двухзеркальной системой является антенна *Кассегрена* (рис. 9.23), состоящая из главного параболического и вспомогательного гиперболического зеркал. Вспомогательное зеркало расположено между фокусом и вершиной главного зеркала таким образом, что один из его двух фокусов совпадает с фокальной точкой главного зеркала. Положение второй фокальной точки определяет фокальную точку системы и, следовательно, положение облучателя.

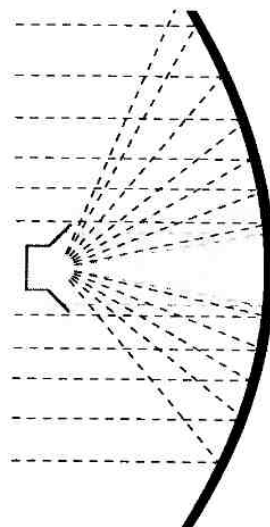


Рис. 9.22. Однозеркальная антенна

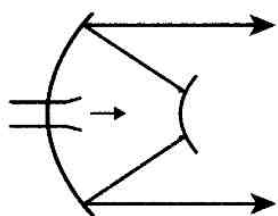


Рис. 9.23. Антенна Кассегрена

Эффективное фокусное расстояние системы Кассегрена можно определить из выражения [81]:

$$F_e = \frac{F_1}{F_2} F = mF,$$

где F_e — эффективное фокусное расстояние антенны Кассегрена; F_1 — расстояние от облучателя до вершины вспомогательного зеркала; F_2 — расстояние от фокальной точки главного зеркала до вершины вспомогательного зеркала; F — фокусное расстояние главного зеркала; m — коэффициент увеличения фокусного расстояния.

Из выражения следует, что эффективное фокусное расстояние может изменяться в пределах от 0 до ∞ в зависимости от положения облучателя и вспомогательного зеркала. На практике величина m обычно берется в пределах от 2 до 6.

Выбор параметров простых конструкций с рупорным облучателем, как правило, основан на компромиссе между размерами раскрыва рупора и размерами вспомогательного зеркала. При этом конструкторы стремятся уравнивать размеры теней, отбрасываемых раскрывами облучателя и вспомогательного зеркала.

Для исключения эффекта затенения апертуры успешно применялись специальные способы. Например, использовалось вспомогательное зеркало из сетки, которое отражало волны одной поляризации и было прозрачно для волн с ортогональной поляризацией, а также сетка вращения поляризации, расположенная на одну четверть длины волны ближе к главному зеркалу. Однако подобные устройства частотно-зависимы и, кроме того, слишком критичны и дороги [81].

Антенна Грегори. В системе Грегори, так же как и в антенне Кассегрена, в качестве главного зеркала используется параболоид, но вспомогательное зеркало в этом случае имеет эллиптическую, а не гиперболическую поверхность. Так как оба фокуса эллипса расположены с одной стороны вспомогательного зеркала, последнее располагают дальше фокальной точки главного зеркала. Из-за такого перемещения вспомогательного зеркала система Грегори менее компактна по сравнению с кассегреновской и потому реже используется на практике.

Перископическая антенна. В перископической антенной системе два зеркала используются в целях изменения направления распространения электромагнитной энергии (рис. 9.24). По этой причине такая система не соответствует приведенному выше определению двухзеркальных антенн. Однако она также заслуживает рассмотрения в целях полноты изложения. Основное применение перископические системы находят в случаях, когда выходной излучатель требуется поднять на большую высоту, чтобы избежать влияния окружающих предметов, не прибегая к использованию длинных линий передачи от наземного электронного оборудования.

В качестве наземного отражателя может использоваться любая антенна с большим усилением. Обычно применяют параболоид с вертикальным направлением излучения. Второй антенной служит либо параболическое зеркало, фокальная точка которого совпадает с наземной антенной, либо плоский рефлектор, переизлучающий энергию в нужном направлении. Характеристики такой системы зависят от многих факторов, в том числе от расстояния между зеркалами.

Типы облучателей. Согласно [81] количество различных типов облучателей, применяемых в зеркальных антеннах СВЧ, превышает 50. Здесь нет возможности остановиться на всех подробно, тем более что многие из них имеют весьма специальное и ограниченное применение. По этой причине далее будут рассмотрены лишь некоторые наиболее распространенные конструкции, среди которых в первую очередь заслуживает рассмотрения рупорный облучатель. Сочетание превосходных характеристик по импедансу, поляризации и допустимой мощности с простотой конструкции привело к широкому применению этого типа облучателей. Рупорный облучатель, как и полноценная рупорная антенна, может иметь пирамидальную, секториальную или коническую конфигурации. Выбор конструкции позволяет получить нужный тип поляризации. Для формирования специальной диаграммы направленности или коррекции aberrаций могут применяться многорупорные системы облучателей.

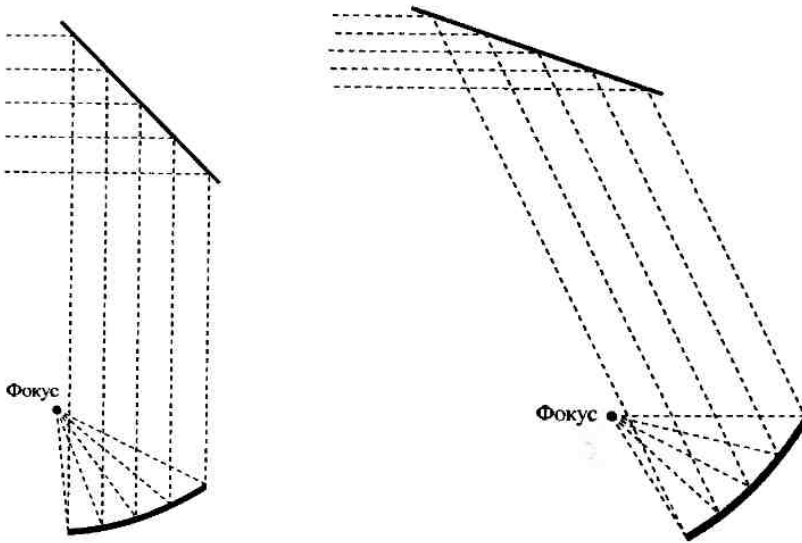


Рис. 9.24. Перископические антенны

Дипольный облучатель менее распространен, чем рупорный. Диаграмма его излучения такова, что для придания ей направленности требуется использование рефлектора. Обычно рефлектор имеет форму плоской пластины или уголкового отражателя. Два скрещенных диполя образуют турникетную антенну. Такая антенна при возбуждении диполей со сдвигом по фазе на 90° позволяет создавать круговую поляризацию. Дипольные облучатели менее универсальны, чем рупорные, однако их можно использовать, когда не требуется максимальная мощность.

Примером простого облучателя для получения круговой поляризации является спираль. В приемной аппаратуре, где ограничения по допустимой мощности в антенне не имеют значения, спираль может работать с хорошими импедансными характеристиками в диапазоне шириной с октаву. Для получения поляризационных характеристик такой облучатель не нуждается в каких-либо специальных фидерных устройствах.

В качестве облучателей в аппаратуре связи типа «точка-точка» широко применяют частотно-независимые структуры в виде плоской и конической спиралей или логопериодические устройства. Данные структуры имеют предельно широкополосные характеристики. Как и в случае простой спирали, более сложные спиральные конфигурации порождают круговую поляризацию без каких-либо специальных фидерных устройств. Недостаток логопериодических структур и конической спирали связан с тем фактом, что положение их фазового центра зависит от частоты. Вследствие этого на диаграмму зеркальной антенной системы, работающей в широком диапазоне частот, будет влиять эффект дефокусировки.

В случае, когда облучатель должен быть линейным, могут использоваться решетки из диполей и рупоров или же решетки из щелевых волноводных антенн.

Возможно создание смешанных (гибридных) систем облучателей, состоящих из структур разных типов. В тех случаях, когда предусматривается работа на нескольких частотах и облучатель одного вида не в состоянии обеспечить нужной широкополосности, применяют рупорный и дипольный облучатели совместно.

В простой зеркальной антенне с точечным фокусом требования к диаграмме направленности облучателя заключаются в обеспечении однородного фазового фронта и создании в апертуре антенны заданного амплитудного распределения. Для многих простых облучателей диаграмму можно аппроксимировать квадратичной функцией до уровня приблизительно -20 дБ.

Выбор типа диаграммы облучателя может диктоваться, помимо амплитудного распределения в апертуре, также и другими соображениями. Во многих малопушующих устройствах условия максимального КНД и отсутствия заметного «переливания» энергии за края зеркала могут быть более значимыми, нежели уровни малых боковых лепестков. Для таких устройств диаграмма облучателя в идеале должна давать равномерное освещение апертуры и спад до нуля на краях.

Однако на практике удастся с помощью облучателей довольно сложной конструкции в лучшем случае получить лишь некоторое приближение к подобию такой диаграммы. Обычно же разработчикам приходится идти на компромисс между величиной КНД, уровнями боковых лепестков и «затекания» энергии, выбирая величину спада амплитуд на краях апертуры равной, например, -15 дБ.

Антенны поверхностных волн представляют собой плоские или стержневые направляющие системы ограниченной длины, вдоль которых могут распространяться поверхностные электромагнитные волны. В качестве направляющих структур применяются диэлектрические пластины, диэлектрические пластины на металлической подложке, плоские либо стержневые ребристые структуры, диэлектрические стержни или трубки и т. д. В зависимости от соотношения между геометрическими размерами направляющей структуры и длиной волны можно получить различную по ширине ДН. Общим свойством этих диаграмм является то, что главный лепесток направлен вдоль оси антенны или несколько отклонен от нее (в несимметричных антеннах). По этой причине антенны поверхностных волн следует отнести к антеннам осевого излучения. Действие таких антенн во многом определяется свойствами поверхностных электромагнитных волн, которые также делятся на E - и H -волны. Соотношения между их полями определяются параметрами диэлектрического волновода, длиной волны и способом возбуждения.

Диэлектрическая антенна (рис. 9.25) [47] состоит из диэлектрического стержня 1 (чаще всего круглого сечения) и возбуждающего устройства, которое пред-

ставляет собой отрезок волновода круглого сечения 2. Внутри волновода находится несимметричный вибратор, подключенный к коаксиальному кабелю 3. В качестве материала для изготовления стержней используется высокочастотный диэлектрик (полистирол, фиберглас и т. п.). Одним из основных требований к материалу являются малые потери.

В отрезке волновода 2 с помощью несимметричного вибратора возбуждается волна типа H_{11} . В диэлектрическом стержне эта волна превращается в волну смешанного типа HE_{11} . Подобного типа волны могут распространяться в диэлектрическом волноводе независимо от соотношения между диаметром стержня, длиной волны и показателем преломления диэлектрика. Однако, как показывает опыт, оптимальные размеры площади поперечного сечения диэлектрического стержня, исходя из условий возбуждения и необходимого замедления, должны лежать в пределах [47]:

$$0,13 \frac{\lambda^2}{N^2 - 1} \leq S \leq 0,28 \frac{\lambda^2}{N^2 - 1},$$

где S — площадь поперечного сечения стержня; λ — длина волны в свободном пространстве; N — показатель преломления диэлектрика.

Если диэлектрический волновод ограничить по длине, то набегающая к его концу волна будет в основном излучаться в окружающее пространство, но частично будет и отражаться, распространяясь в сторону возбуждающего устройства. Наличие отраженной волны связано с резким переходом от условий распространения вдоль диэлектрического стержня с замедленной скоростью к условиям распространения в окружающем пространстве. Для уменьшения отражения от конца стержня его делают конусообразным с закруглением на конце. В этом случае по мере приближения к краю стержня фазовая скорость распространения волны увеличивается, чем обеспечивается плавный переход от распространения вдоль диэлектрического стержня к излучению в воздушном пространстве. Вместе с уменьшением отражения наблюдается некоторое сужение ДН и понижение уровня боковых лепестков. Диэлектрические стержневые антенны в системах связи применяют, как правило, на частотах от 2 ГГц и выше.

К современным антенным устройствам широкополосной связи предъявляются весьма жесткие требования по электродинамическим характеристикам, а также габаритам, массе, стоимости, технологичности и т. д. Из-за необходимости создания невыступающих, низкопрофильных, а также простых и технологичных антенн широкого назначения возник большой интерес к микрополосковым и печатным антеннам СВЧ. Использование планарной технологии обеспечивает высокую воспроизводимость размеров антенн, технологичность, низкую стоимость, малые габариты и массу. Методами планарной технологии могут быть выполнены не только излучатели, но и линии передачи, согласующие элементы, фазовращатели, переключатели и пр.

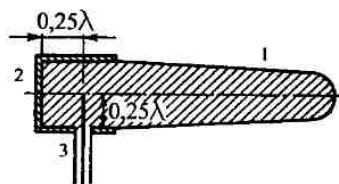


Рис. 9.25. Диэлектрическая стержневая антенна

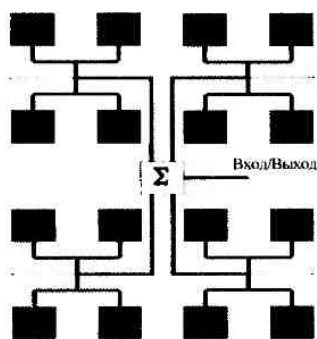


Рис. 9.26. Печатная антенная решетка

Под плоскостными (планарными) антеннами СВЧ, согласно [11], следует понимать как одиночные микрополосковые и печатные излучатели, так и образованные из них антенные решетки (рис. 9.26). Антенны СВЧ могут излучать волны линейной, круговой и эллиптической поляризации, обеспечивать работу в многочастотных режимах.

Элементы печатных излучающих структур могут иметь самую разнообразную геометрическую форму: прямоугольную, круглую, треугольную, эллиптическую, кольцевую и др. Эти структуры способны работать как в резонансном, так и в нерезонансном режиме с возбуждением от полосковой линии, коаксиального кабеля, за счет электромагнитной связи.

Недостатками планарных антенн являются узкополосность, малые предельно допустимые мощности, трудность конструирования перестраиваемых устройств и изменения параметров печатных элементов.

Антенной решеткой (АР) называется система идентичных излучающих элементов, расположенных по определенному закону.

Такие конструкции используются для получения высокой направленности излучения с помощью системы слабонаправленных антенн типа вибраторов, щелей, открытых концов волноводов, возбуждаемых токами с заданным соотношением амплитуд и фаз. В этом случае общая направленность, особенно при большом числе излучателей, определяется в основном габаритными размерами всей системы и в гораздо меньшей степени — индивидуальными направленными свойствами отдельных излучателей.

Как правило, АР создаются для получения диаграмм направленности с главным лучом малой ширины, угловое положение которого можно менять тем или иным способом (механическим, электромеханическим или электрическим). Целесообразность использования АР становится особенно ясной тогда, когда требуется перемещать один или несколько главных лучей в пространстве достаточно быстро и в порядке, определяемом перемещением абонентских терминалов.

В зависимости от расположения антенных элементов различают *линейные, поверхностные* и *объемные* решетки, среди которых наибольшее распространение получили *прямолинейные* и *плоские* АР. Иногда излучающие элементы располагаются по дуге окружности или на криволинейных поверхностях, совпадающих с формой объекта, на котором расположена АР. Такие АР называются *конформными*.

Простейшей является линейная АР, в которой излучающие элементы расположены вдоль прямой, называемой *осью решетки*, на равных расстояниях друг от друга (*эквидистантная АР*). Расстояние d между фазовыми центрами парциальных излучателей называют *шагом решетки*. Линейная АР помимо самостоятельного значения зачастую является основой при анализе других типов АР. В современных базовых станциях сотовой связи второго поколения для обеспечения круговой работы используются, как правило, три вертикальные решетки излучателей (рис. 9.27).

Принцип работы всякой решетки излучателей сводится к созданию таких фазовых сдвигов в питающих токах (при работе на передаче), при которых волны, формируемые антенными элементами, в некотором направлении складываются синфазно в дальней зоне. При работе на прием выходные сигналы на элементах решетки изменяют по фазе таким образом, чтобы они оказались синфазными на входе суммирующего устройства для заданного направления приема.

Рассмотрим принцип действия АР на примере простейшей системы ненаправленных излучателей, расположенных на одинаковых расстояниях друг от друга вдоль одной линии эквидистантной АР. Предположим, что на решетку падает плоская волна, направление прихода которой перпендикулярно раскрытию антенны.

Поле в месте расположения элементов одинаковое, следовательно, напряжения на выходах антенных элементов также будут иметь одинаковые амплитуды и фазы, а напряжение на выходе сумматора будет постоянным.

При изменении направления прихода волны амплитуда поля по раскрытию антенны не меняется, но появляются фазовые сдвиги, величина которых зависит от угловой координаты источника излучения, расстояния между элементами и длины волны. Межэлементный набег фазы определяется величиной $2\pi d(\sin\theta)/\lambda$, где θ — угол между направлением на источник излучения и нормалью к решетке.

Для того чтобы АР имела максимальную направленность по углу θ , необходимо либо механически повернуть ее, либо ввести с помощью фазовращателей фазовые сдвиги в напряжения, подаваемые на сумматор.

ДН k -го антенного элемента $f_k(\theta)$ эквидистантной линейной антенной решетки после синфазного суммирования напряжений по выходам R излучателей имеет вид:

$$f_k(\theta) = F_k(\theta) \frac{\sin \left[\frac{R\pi d}{\lambda} (\sin \theta - \sin \theta_k) \right]}{\sin \left[\frac{\pi d}{\lambda} (\sin \theta - \sin \theta_k) \right]},$$

где $F_k(\theta)$ — ДН одиночного элемента, θ_k — направление максимума ДН.

Если ДН излучателей идентичны и изотропны (не зависят от направления излучения), то выражение упростится:

$$f_k(\theta) = \frac{\sin \left[\frac{R\pi d}{\lambda} (\sin \theta - \sin \theta_k) \right]}{\sin \left[\frac{\pi d}{\lambda} (\sin \theta - \sin \theta_k) \right]}.$$

Ширина диаграммы направленности на уровне половинной мощности определяется выражением:

$$\theta_{0,5} \approx \frac{0,888\lambda}{Rd \cos \theta_k}.$$

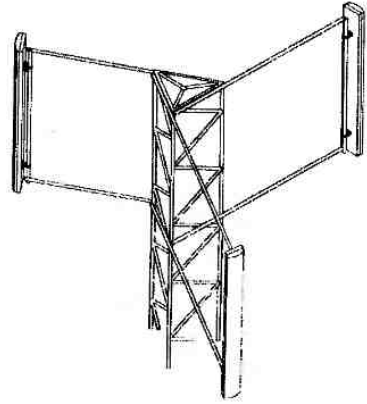


Рис. 9.27. Вертикальные линейные решетки излучателей антенной системы базовой станции

Таким образом, ширина ДН зависит от размера линейной решетки Rd , длины волны λ и величины отклонения луча θ . По мере отклонения луча от направления нормали к решетке происходит его расширение в $1/\cos\theta_k$ раз.

Особенностью антенной решетки, как дискретной системы излучателей, является повторяемость максимумов ДН. Заслуживает внимания случай, когда расстояние между соседними излучателями больше половины длины волны. При этом в диаграмме направленности могут появиться боковые лепестки, максимумы которых равны по амплитуде максимуму главного лепестка. Они возникают в тех направлениях, для которых разность фаз полей соседних излучателей равна $2\pi n$, где $n = \pm 1, \pm 2, \dots$ ($n = 0$ соответствует главному максимуму). Эти боковые лепестки называют дифракционными максимумами. Направления дифракционных максимумов и их количество зависят от длины волны, расстояния между соседними излучателями и направления главного максимума. Для борьбы с дифракционными лепестками ДН используют направленные свойства антенного элемента решетки $F_k(\theta)$. При этом выбирают излучатели с такой ДН $F_k(\theta)$, чтобы интенсивность дифракционных максимумов в значительной степени ослаблялась.

Если межэлементное расстояние меньше или равно половине длины волны излучения, то дифракционные максимумы отсутствуют при любых положениях главного лепестка.

Электрическое управление положением главного луча может осуществляться фазовым либо частотным методом. При фазовом методе управления используются фазовращатели, с помощью которых обеспечивается изменение сдвига по фазе между токами в излучающих элементах. Изменение фазового сдвига может быть плавным или дискретным. В тех случаях, когда требуются большие скорости изменения фазы, применяются электрические фазовращатели на полупроводниковых диодах или ферритах. Чаще всего используются диоды с $p-i-n$ -полупроводниковой структурой.

При частотном управлении лучом меняется частота питающего решетку генератора. Это приводит к изменению сдвига фаз между токами в излучателях решетки, в результате которого меняется положение антенного луча. Частотное управление является наиболее простым с точки зрения технического выполнения, но требует перестройки частоты передатчика в сравнительно широких пределах (не менее 6% от несущей частоты). Поэтому в системах связи оно не используется.

Фазированные антенные решетки (ФАР) — это антенные системы, состоящие из элементов с независимым управлением фазой. В решетке подобного типа используется самый распространенный способ электрического сканирования — фазовый.

Наиболее широко применяемым типом ФАР в системах связи являются линейные и плоские решетки. В большинстве случаев плоская ФАР состоит из идентичных излучателей, расположенных в узлах плоской координатной сетки с прямоугольными или треугольными (гексагональными) ячейками.

Различают активные и пассивные ФАР. В активных ФАР (АФАР) [48] каждый элемент возбуждается от отдельного генератора или усилителя мощности.

В пассивных ФАР все излучатели запитываются от общего генератора или работают на общий приемник. Различают ФАР с фидерным и пространственным питанием (пространственное питание состоит в том, что энергия на антенную решетку поступает от облучателя, например рупора).

Фидерное питание может осуществляться по последовательной или параллельной схеме. Применяются также их комбинации.

При большом количестве излучателей фидерные схемы питания отличаются сложностью и громоздкостью и в основном находят применение в диапазонах дециметровых и более длинных волн. Однако в системах связи они стали использоваться и в сантиметровом диапазоне, где прежде предпочтение отдавалось пространственной схеме питания.

9.6 Фрактальные антенны

Среди множества антенных конструкций, используемых сегодня в широкополосной связи, фрактальные антенны принципиально отличаются от известных решений. Фактически, вся предшествовавшая история эволюции антенн базировалась на использовании достижений евклидовой геометрии, оперирующей объектами целочисленной размерности — линия, отрезок, плоскость, круг, эллипс, параболоид и т. п.

Как известно, фракталы — это геометрические формы, рекурсивно повторяющиеся в увеличивающемся либо уменьшающемся масштабе, главным отличием которых является дробная размерность. Такие геометрические объекты были известны задолго до Бенуа Мандельброта, но именно он первым указал на родство их с природно-естественными формами, систематизировал и обобщил этот класс объектов, введя в 1975 году понятие фрактала (от лат. fractus — «расколотый камень») [175]. Открытие фракталов привело к революции не только в геометрии, но и в физике, химии, биологии. Фрактальные технологии нашли применение в области компрессии данных, фильтрации сигналов, синтеза трехмерных компьютерных моделей природных ландшафтов. Неудивительно, что во всеобщий фрактальный бум оказалась вовлеченной и теория антенн.

Первые публикации, рассматривавшие электродинамику фрактальных структур, появились еще в 80-е годы XX века. Начало же практическому использованию фрактального направления в антенной технике более 10 лет назад положил американский инженер Натан Коэн [8.18], ныне профессор Бостонского университета и главный технический инспектор компании «Fractal Antenna Systems». Проживая в центре Бостона, чтобы обойти запрет городских властей на установку наружных антенн, он решил замаскировать антенну любительской радиостанции под декоративную фигуру из алюминиевой фольги. За основу им была взята известная в геометрии кривая (рис. 9.28) [35, 158, 233], описание которой в 1904 году предложил шведский математик Нильс Фабиан Хельге фон Кох (1870–1924).

9.6.1 Излучатель на основе фрактала Коха

Процесс построения кривой Коха состоит в следующем (рис. 9.29). Исходный отрезок длиной z делится на три равные части, и вместо центральной подставляется излом в виде угла равностороннего треугольника со стороной $z/3$. Таким образом, получается трехсегментная структура (образующий шаблон), центральный элемент которой замещен треугольником без основания. Далее процесс повторяется над каждым сегментом в отдельности: во второй итерации — на от-

резках $z/3$ строятся треугольники со сторонами $z/9$, на них — треугольники со сторонами $z/27$ (третья итерация) и т. д. Каждый шаг синтеза увеличивает длину результирующей кривой в $4/3$ раза по отношению к первоначальной. Однако при этом габаритная высота фрактала не изменяется. Поэтому, если бы процесс его формирования был выполнен бесконечное количество раз, полученная кривая имела бы бесконечную длину при неизменной высоте. В общем случае длина кривой Коха определяется выражением [35, 158]:

$$L = z \left(\frac{4}{3} \right)^n,$$

где n — число итераций, z — высота образующего шаблона (длина исходного отрезка).

Рис. 9.28. Диполь на основе фрактала Коха



Как и всякий детерминированный фрактал, кривая Коха обладает самоподобием или масштабной инвариантностью: каждый ее элемент в уменьшенном виде строго повторяет исходную кривую — шаблон. Во фракталах со случайной структурой такой эффект отсутствует.

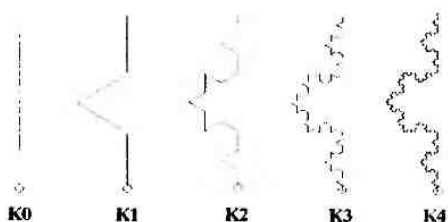


Рис. 9.29. Итерации формирования кривой Коха

Прообразы современных фрактальных технологий появились в антенной технике еще в середине 60-х годов прошлого века, когда были созданы логопериодические и спиральные антенны. Будучи образованными из самоподобных фрагментов, эти антенны, с позиций сегодняшнего дня, также можно обобщить на класс фрактальных, хотя в строгом математическом смысле они являются лишь фракталами первого рода и не имели никакого отношения к фрактальной

геометрии на момент разработки [233]. Быть может, на этом фоне новшество Козэна так и осталось бы декоративной причудой, если бы не выяснилось, что фигура Коха, полученная в результате трех итераций и наклейки на небольшой лист бумаги, обладает пространственно-частотной избирательностью, не уступающей обычным антеннам.

Теория фрактальных антенн пока находится на этапе становления. В основном исследователи экспериментальным путем, методом проб и ошибок, пытаются применить известные в геометрии и алгебре фракталы к антенным конструкциям. В результате многочисленных экспериментов установлено, что подобного

типа антенны позволяют получить практически тот же коэффициент усиления, что и обычные, при меньших габаритных размерах, что особенно важно для мобильных приложений. Например, на рис. 9.30 показаны относительные высоты нескольких итераций фрактала Коха для одной и той же резонансной частоты в сравнении с плечом обычного полуволнового диполя [158].

Эффект миниатюризации антенн наиболее существенно проявляется лишь для нескольких первых итераций фрактала (обычно 5–6), асимптотически приближаясь к некоторому пределу. Что касается входного сопротивления, то на рис. 9.31 приведены расчетные значения для его активной и реактивной составляющей, полученные на основе метода моментов для пяти первых итераций кривой Коха (K0–K4) [158].

Резонансные частоты монополя, как известно, приходятся на нули реактивных сопротивлений. Первый такой нуль, расположенный на восходящей ветви импедансной кривой рис. 9.31, б, соответствует основной резонансной волне или первой гармонике. Слева от этой точки реактивное сопротивление носит емкостной характер (отрицательная величина), а справа — индуктивный. В указанной точке нулевой реактивности фрактальный монополь ведет себя как последовательный резонансный контур. Нули же на нисходящей ветви импедансной кривой (рис. 9.31, б)

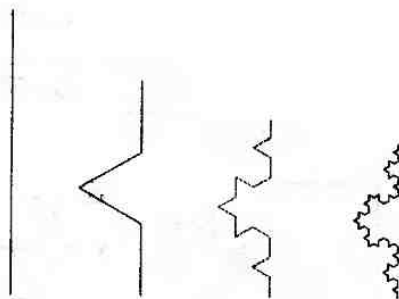


Рис. 9.30. Сравнительные размеры фракталов Коха и четвертьволнового монополя

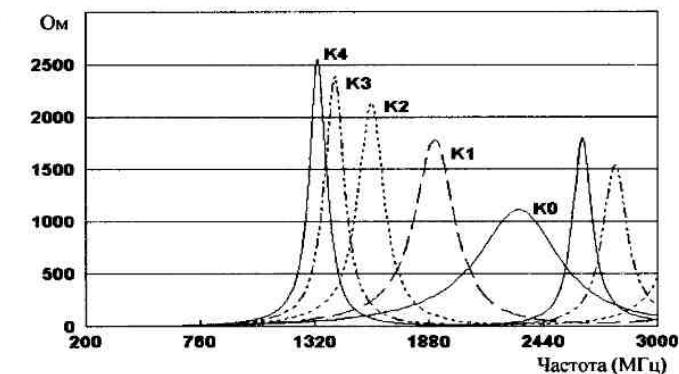
приходятся на параллельные резонансы (эквивалентная электрическая схема монополя соответствует параллельному резонансному контуру), для которых характерны максимумы значений активного сопротивления (рис. 9.31, а). С увеличением частоты сигналов ход кривых качественно повторяется, с той лишь разницей, что пиковые значения активного и реактивного сопротивлений постепенно убывают по величине. В табл. 9.3 приведено сравнение резонансных частот итераций фрактала K1–K4 и обычного четвертьволнового вибратора (K0) для той же самой длины его проводника диаметром 0,2 см, что и в антенне Коха. Следует обратить внимание, что с увеличением диаметра провода указанные резонансные частоты рекурсивного монополя Коха увеличиваются.

Для сопоставления электрических свойств антенны Коха с другими излучателями периодической структуры можно привести экспериментальные данные, полученные в [128] в отношении кривой Коха, четырех итераций меандра (рис. 9.32) и спиральной антенны, варианты которой представляют собой $2\frac{1}{4}$, $2\frac{3}{4}$ и $3\frac{1}{8}$ витков провода (рис. 9.33) при общем диаметре спирали около 1 см.

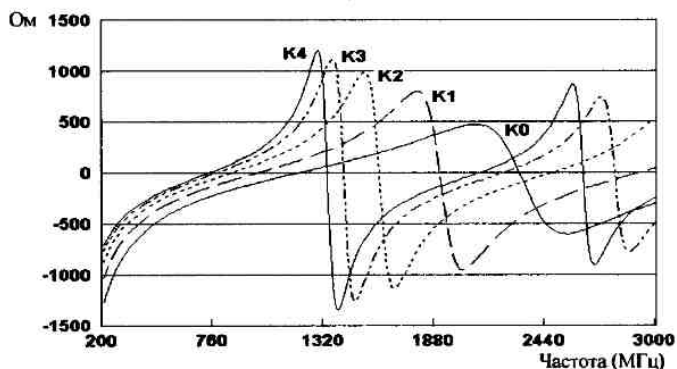
Экспериментально снятые зависимости сопротивлений указанных антенн на примере итераций K3, M3 и H3 приведены на рис. 9.34 для случая, когда все излучатели имели одинаковую высоту 15,24 см (данные габариты были взяты для упрощения изготовления антенн). Нетрудно заметить, что меандр (M3) и спиральная антенна (H3) имеют более сильный спад резонансных частот при

одинаковой высоте излучателей. Это позволяет в случае использования МЗ- и НЗ-антенн при одинаковой резонансной частоте несколько снизить общую высоту антенны в сравнении с кривой Коха.

Рис. 9.31. Расчетные значения активного (а) и реактивного (б) сопротивлений монополей Коха



а)



б)

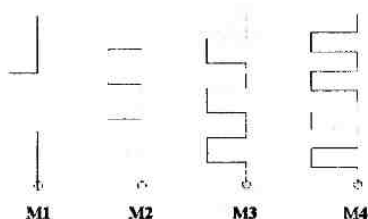


Рис. 9.32. Монополи в виде меандра

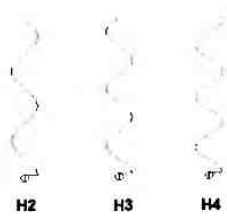


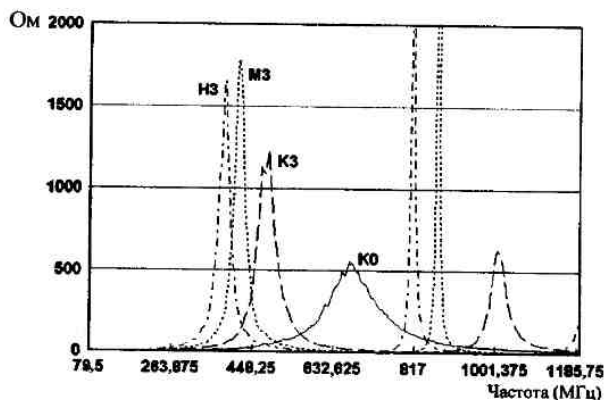
Рис. 9.33. Спиральные монополи

Все сопоставленные измерители обладают многочастотными свойствами, что проявляется в наличии периодических резонансов на графиках импедансов. Однако для многодиапазонных приложений более всего пригоден фрактал Коха, у которого с ростом частоты пики реактивных и активных сопротивлений уменьшаются, тогда как у меандра и спирали они растут по величине.

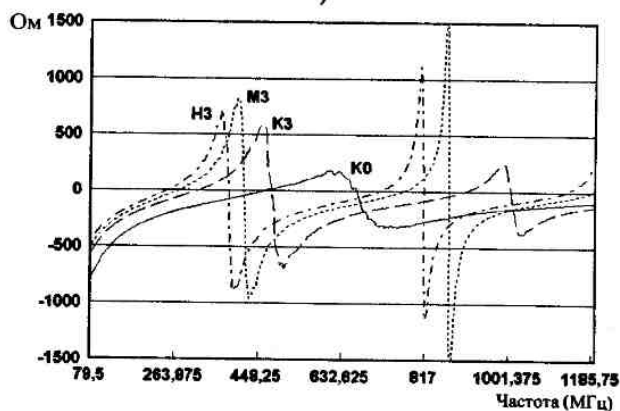
Таблица 9.3. Сравнение резонансных частот итераций фрактала K1-K4 и обычного четвертьволнового вибратора (K0)

Антенна	Высота антенны, см	Общая длина проводника, см	Резонансная частота, МГц	Резонансная частота монополя K0 той же длины, МГц
Евклидов монополь (K0)	6	6	1195,2	1195,2
K1	6	8	987,2	898,5
K2	6	10,67	853,2	675,2
K3	6	14,22	777,8	507,4
K4	6	18,96	740,0	381,2

Рис. 9.34. Сравнение активных (а) и реактивных (б) сопротивлений монополей в виде фрактала Коха, менадра и спирали



а)



б)

9.6.2 Диполь на основе фрактала Минковского

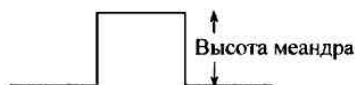


Рис. 9.35. Образующий меандр петли Минковского

Весьма близким по своим свойствам к антенне на основе фрактала Коха является диполь, сформированный по закону ломаной Германа Минковского (1864–1909) [35, 158]. При построении этой антенны вместо системы треугольников на прямой формируются меандры убывающих размеров. При первом изгибе прямолинейного диполя в виде «прямоугольного солитона» (рис. 9.35) его усиление возрастает. После-

дующие итерации практически не изменяют коэффициент усиления, но диапазон рабочих частот диполя расширяется, а сама антенна становится гораздо компактнее. Как и в случае кривой Коха, эффективными являются лишь первые 5–6 шагов: чтобы сгибать провод дальше, придется уменьшить его диаметр, а это повысит сопротивление антенны и приведет к потере усиления.

9.6.3 Древовидные диполи

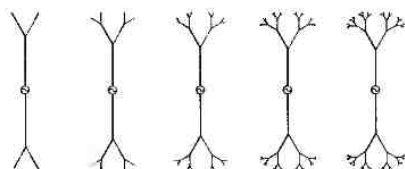


Рис. 9.36. Двумерные диполи в виде рекурсивных деревьев

Другой тип фрактала, который может использоваться в качестве диполя, — двумерное рекурсивное дерево, чьи варианты геометрии приведены на рис. 9.36 [158]. Фрактал генерируется из простого монополя путем последовательного разбиения его вершин на две ветви под заданным углом (до 60°). Полная электрическая длина такой антенны может быть определена как самая короткая длина от ядра фрактала к

любому его концу. Каждая новая итерация увеличивает количество проводящих путей на краях антенны и при неизменной высоте дерева понижает резонансную частоту. Характеристика направленности древовидного диполя в дальней зоне очень близка к таковой для прямого диполя.



Рис. 9.37. Итерации трехмерных диполей в виде фрактальных деревьев

Дальнейшим развитием антенн типа «двумерное дерево» являются «трехмерные деревья», получаемые, например, путем последовательного разбиения вершин прямого монополя на четверки ветвей в двух ортогональных плоскостях. Соотношения размеров каждой из ветвей для разных итераций при угле разбиения 60° приведены в табл. 9.4 [158]. Вид первых пяти итераций фрактала показан на рис. 9.37.

Как видно, наряду с минимизацией габаритов при использовании такого типа антенн становятся излишними дополнительные меры по маскировке их под есте-

ственные деревья, как это нередко делалось до сих пор в случае традиционных антенных систем [94].

Таблица 9.4. Соотношения размеров ветвей для разных итераций при угле разбегания 60°

	Номер итерации					
	0	1	2	3	4	5
Относительные размеры ветвей	1	1/3	1/7	1/15	1/31	1/63
		2/3	2/7	2/15	2/31	2/63
			4/7	4/15	4/31	4/63
				8/15	8/31	8/63
					16/31	16/63
						32/63

В целом выигрыш от миниатюризации дипольных структур существенно зависит от типа используемого при их формировании фрактала. Обладая наибольшей компактностью, трехмерные фрактальные деревья имеют значительно меньшее входное сопротивление по сравнению с двумерными аналогами. На рис. 9.38 для сравнения приведены зависимости резонансной частоты от номера итерации для кривой Коха, двумерного и трехмерного деревьев [158], а на рис. 9.39 — аналогичные графики для входного сопротивления. Таким образом, при ориентации на трехмерную фрактальную антенну древовидного типа необходимо искать разумный компромисс между достижимой компактностью антенны и снижением величины ее входного сопротивления, приводящим к проблемам в согласовании с питающим фидером.

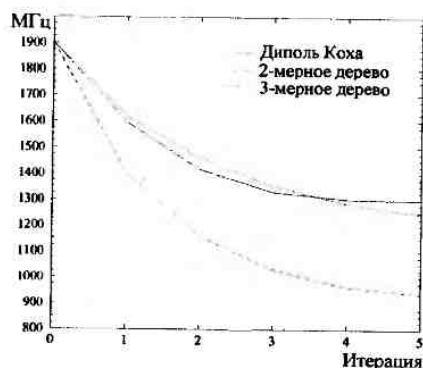


Рис. 9.38. Зависимости резонансных частот от номера итерации фрактала

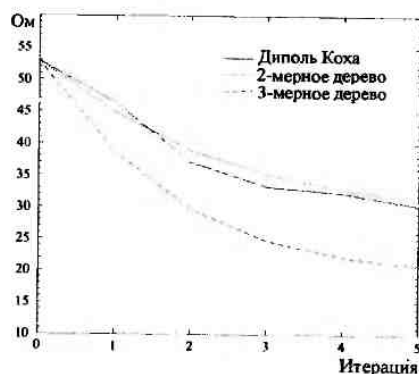


Рис. 9.39. Зависимости сопротивлений от номера итерации фрактала

Все рассмотренные дипольные антенны являются довольно узкополосными. К примеру, ширина полосы частот на уровне 3 дБ у обычного диполя составляет 2,4% от несущей, для кривой Коха в 5-й итерации этот параметр возрастает

до 3,1%, использование двумерного древовидного фрактала 5-й итерации позволяет расширить полосу приема до 4,2% и лишь в варианте трехмерного дерева — до 12,7%. Однако в последнем случае на резонансной частоте имеют место указанные потери входного сопротивления порядка 7,75 дБ [158].

В настоящее время еще не изучены свойства антенн, представляющих собой комбинацию нескольких различных фракталов, например Коха и трехмерного дерева. Не исключено, что рекурсия Минковского либо Коха в древовидных ветвях может уменьшить потери входного сопротивления, что сделает трехмерные фрактальные антенны еще более привлекательными. Нуждаются в более детальном изучении и поляризационные свойства фрактальных антенн.

9.6.4 Рамочные антенны на основе фрактальных петель Коха и Минковского. Крест Минковского

Помимо дипольных структур в качестве узкополосных антенн могут использоваться фрактальные рамки, построенные на основе рассмотренных выше кривых Коха и Минковского (рис. 9.40). Как известно, традиционные рамочные антенны при малых размерах имеют низкое входное сопротивление, что усложняет их согласование с 50-омным питающим фидером. Фрактальные же технологии позволяют заметно увеличить сопротивление рамочной антенны на частотах, ниже резонансной, упрощая решение задачи ее согласования для нужд радиопеленгации и частотного мониторинга, а также апертурную эффективность.



Рис. 9.40. Фрактальные рамки на основе кривых Минковского (слева) и Коха (справа)

Чтобы проиллюстрировать выигрыш, предоставляемый фрактальным подходом, сопоставим круглую рамку и «петлю Коха» четвертой итерации, относительные размеры которых приведены на рис. 9.41. Основные соотношения для расчетов следующие [158]:

- а) площадь, ограниченная фракталом Коха четвертой итерации:

$$S_{\text{Коха}} = 2,05r^2,$$

- б) периметр петли Коха

$$P_{\text{Коха}} = 3\sqrt{3}r \left(\frac{4}{3}\right)^n,$$

где n — номер итерации (при $n = 4$ · $P_{\text{Коха}} = 16,42r$).

Следует отметить, что площадь, охваченная данной разновидностью фрактала, на 35% меньше площади описанного круга. В то же время фрактальный периметр превышает периметр круговой петли в 2,614 раза. Согласно [158] такие соотношения геометрических параметров приводит к тому, что при диаметре проволоки $1,4 \times 10^{-5} \lambda$ и периметре петли Коха $0,68 \lambda$ ее входное сопротивление равно 26,65 Ом. В то же время круговая рамка на основе описанного вокруг этого фрактала окружности имеет периметр $0,26 \lambda$ и входное сопротивление всего лишь 1,17 Ом. Незначительная реактивность петли Коха теоретически может быть компенсирована добавлением параллельной контуры емкости.

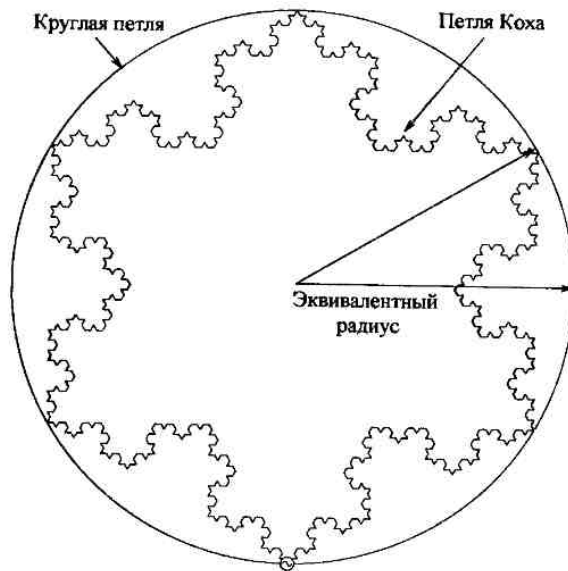


Рис. 9.41. Относительные размеры круглой рамки и вписанной в нее петли Коха

Для рассмотренных размеров рамок электрическая длина петли Коха оказывается в 32,2 раза больше физической апертуры, охватываемой ею площади, тогда как для круговой рамки это соотношение падает до 22,1 раза. Что касается характеристик направленности сопоставленных рамочных антенн, то они очень похожи, но не тождественны: из-за большого периметра ДН фрактальной антенны имеет более выраженный многолепестковый характер.

В одном из последних изобретений Натана Козна [145] предлагается размещать внутри мобильного телефона несколько фрактальных антенн Минковского (рис. 9.42), коммутируемых на вход приемного устройства в зависимости от ориентации их диаграмм направленности на базовую станцию. При этом печатная плата телефона заметно экранирует излучение антенны в направлении головы пользователя. Кстати, в этом же изобретении предложен вариант антенны с вертикальной ориентацией диагонали креста Минковского, что поразительно напоминает внешний вид крестов православных храмов (рис. 9.42).

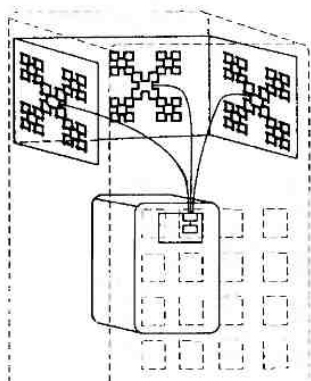


Рис. 9.42. Петли Минковского в качестве антенн мобильного телефона

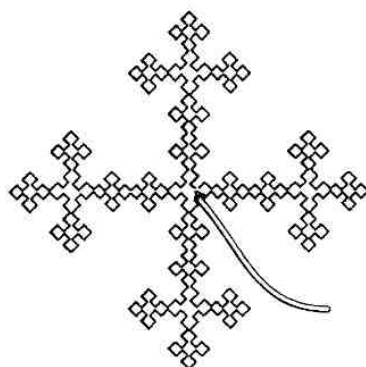


Рис. 9.43. Антенна на основе креста Минковского

Печатная фрактальная антенна в зависимости от ее толщины и диэлектрической постоянной может излучать поверхностные волны. Поэтому, если такой режим является нежелательным, необходимо более тщательно подходить к выбору параметров микрополосковой антенны. Для сокращения длины согласующего трансформатора его также имеет смысл изготавливать на фрактальной основе. Кроме того, фрактальные рамочные и дипольные антенны могут использоваться в сочетании с фрактальным экраном, например на основе ломаной Давида Гильберта (1862–1943) (рис. 9.44) или древовидного фрактала, как это предложено в патенте США № 6140975 [145] (см. рис. 9.44).

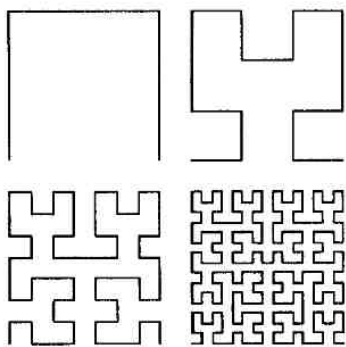


Рис. 9.44. Фракталы Гильберта

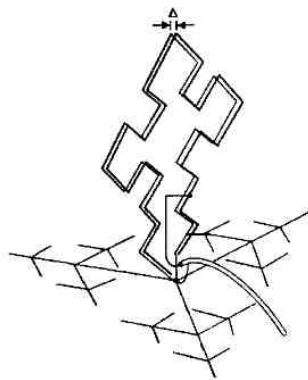


Рис. 9.45. Петля Минковского над древовидным экраном

При таком комплексном применении фрактальных технологий можно сократить расход материалов и вес антенных систем, особенно если излучатели объединены в антенную решетку. В последнем случае миниатюризация антенных элементов на фрактальной основе позволяет существенно снизить их взаимное влияние за счет увеличения межэлементного зазора. В качестве примера на рис. 9.46

проиллюстрированы линейная решетка Дольфа–Чебышева с пятью квадратными излучателями, расположенными с шагом в $0,3\lambda$, и аналогичная ей решетка с элементами Минковского, переход к которой позволяет увеличить зазор между элементами в 5 раз. При этом благодаря ослаблению взаимного влияния фрактальная антенная решетка, согласно [158], обеспечивает уровень боковых лепестков ниже $-26,8$ дБ в секторе сканирования 150° относительно нормали. Расчетные значения уровня максимального бокового лепестка ДН решетки в предположении идеальной компенсации взаимного влияния при этом составляли -30 дБ. Квадратные же элементы в такой решетке приводят к росту первого бокового лепестка до уровня $-4,3$ дБ.

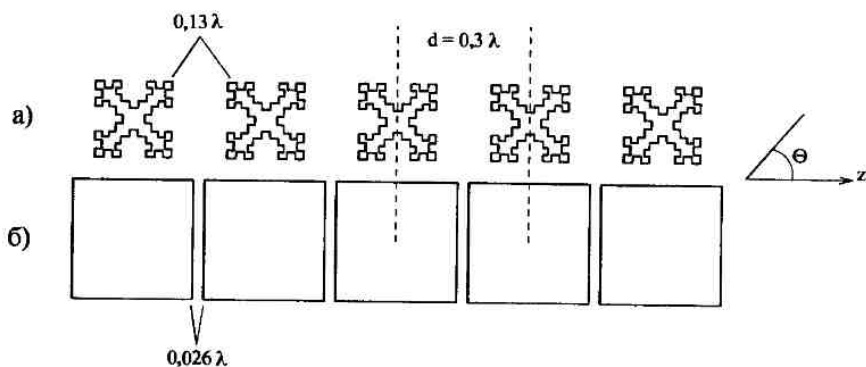


Рис. 9.46. Сравнительные габариты антенной решетки с фрактальными (а) и квадратными (б) элементами

Другой эффект, получаемый вследствие компактности фрактальных излучателей в антенных решетках, состоит в возможности более плотной упаковки антенных элементов в интересах расширения сектора сканирования. Например, в габаритах пятиэлементной эквидистантной решетки Дольфа–Чебышева с полуволновым межосевым интервалом размещения квадратных излучателей можно расположить семь элементов Минковского с той же величиной коэффициентов взаимного влияния, как показано на рис. 9.47.

9.6.5 Монополь, ковер и куб Серпинского. Проволочные трехмерные фракталы

В дополнение к рассмотренным классам сравнительно узкополосных фрактальных конструкций следует остановиться на еще одном типе антенн, в которых самоподобие геометрии приводит к ярко выраженным широкодиапазонным свойствам. Характерным примером в этом ряду является решетка Серпинского [35, 158, 233] (рис. 9.48), на котором впервые были изучены многодиапазонные свойства фрактальных антенн [187]. Данная фигура названа по имени польского математика Вацлава Серпинского (1882–1969), который в 1916 году впервые исследовал ее свойства.

Процесс формирования такого фрактала на первом шаге состоит в удалении из исходного треугольника центрального треугольного сегмента с вдвое меньшей высотой. В образующихся в результате новых треугольниках на втором шаге вновь удаляются центральные части (см. рис. 9.49), и далее эта процедура последовательно повторяется требуемое количество раз.

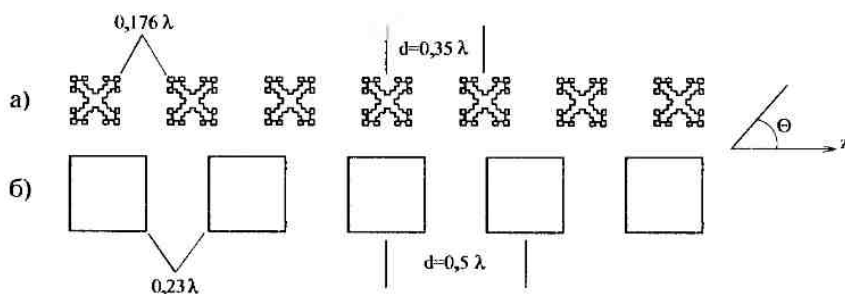


Рис. 9.47. Фрактальный подход позволяет плотнее размещать антенные элементы при фиксированном их взаимном влиянии

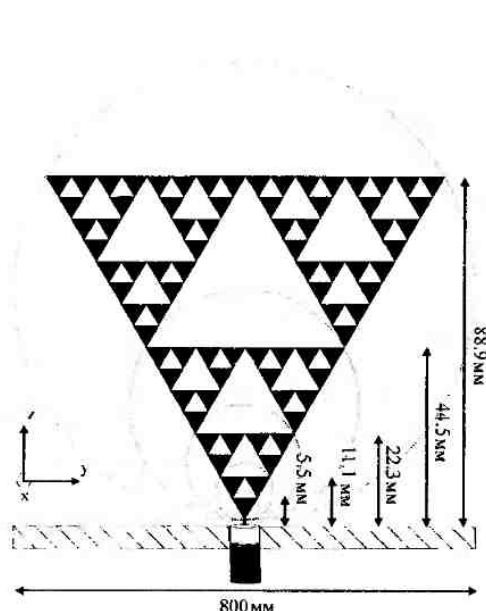


Рис. 9.48. Монополь в виде решетки Серпинского, показанный, например, на рис. 9.50.

Свойство самоподобия фрактальных структур обусловило высокую повторяемость их электромагнитных параметров в различных частотных диапазонах. Речь идет о входном сопротивлении, распределении плотности токов, диаграммах направленности излучения и других параметрах антенн. Все это существенно упрощает их компьютерный расчет и проектирование.

При создании решетки Серпинского для обеспечения протекания тока от одного треугольника к другому края отверстий скругляют. Пример распределения резонансных частот в монополярном варианте такой антенны, приведенном на рис. 9.48, представлен в табл. 9.5 [187]. Вторая и последующие резонансные частоты монополя в виде решетки Серпинского определяются по упрощенной формуле [187]:

$$f_n \approx 0,26 \frac{c}{h} 2^{n-1},$$

где n — номер резонанса ($n \geq 2$), c — скорость света, h — высота наибольшего по размеру треугольного сегмента.

Объединением двух монополей несложно получить диполь Серпинского.

Рис. 9.49. Итерации решета Серпинского

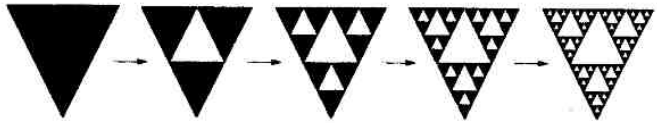


Таблица 9.5. Пример распределения резонансных частот в монопольном варианте решета Серпинского

Номер полосы (n)	f_n , ГГц	Ширина полосы, %	h_n/λ_n
1	0,52	7,15	0,153
2	1,74	9,04	0,258
3	3,51	20,50	0,261
4	6,95	22,00	0,257
5	13,89	25,00	0,255

Вместе с тем конечность размеров, ограниченное число фрактальных итераций при формировании геометрии антенного полотна, влияние цепей питания и симметрирования антенны приводят к появлению погрешностей в реализации электромагнитных свойств фрактальных конструкций. Если соответствующие отклонения не превышают 10%-ного порога, то с ними, как правило, вынуждены мириться [158], поскольку фрактально обусловленные эффекты при этом продолжают иметь место. При большем уровне погрешностей одним из путей их компенсации является намеренное внесение предсказаний за счет нарушения симметрии фрактала [224]. В патенте США [223] предложено использовать решето Серпинского в качестве автомобильной антенны, размещенной на лобовом стекле салона. Выполненная из токопроводящей оптически прозрачной пленки, такая конструкция позволяет при необходимости вести прием во всех известных диапазонах частот FM-вещания, осуществлять радиотелефонную и транкинговую связь, беспроводной доступ к широкополосным сетям.

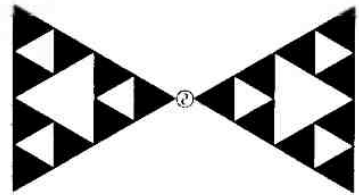


Рис. 9.50. Диполь Серпинского

В сотовых радиотелефонах получила также распространение многодиапазонная фрактальная антенна на основе прямоугольного ковра Серпинского, формируемого в отличие от рассмотренного решета изъятием из исходного квадрата убывающих по величине квадратных сегментов (рис. 9.51). Такая антенна достаточно компактна и может не только устанавливаться на плате, но и напыляться на внутреннюю поверхность корпуса смартфона. Для ее изготовления, как правило, используют фаббер-технологии трехмерного прототипирования, позволяющие прецизионно сформировать миниатюрные антенны любой сложности [91]. Благодаря развитию теории фрактальных антенн серьезное продвижение намечилось и в реализации проектов создания систем на одном чипе с интеграцией на кристалл либо корпус процессора обработки сигналов компактных антенных из-

лучателей. Такие решения довольно перспективны для стандарта HIPERLAN-2 (17,2 ГГц) и других сетей беспроводного доступа.

Как и в случае фрактальных деревьев, рассмотренные многодиапазонные антенны могут обобщаться на трехмерный случай, например, развитием ковра Серпинского является фрактальный куб, а решетка — пирамида.

9.7 SMART-антенны. Цифровые антенные решетки (ЦАР)

9.7.1 Понятие ЦАР. Преимущества ЦАР по сравнению с ФАР

Отличительной особенностью известных концепций систем мобильной связи 3-го и 4-го поколений является использование в базовых станциях технологии цифрового формирования луча. Средством реализации таковой являются цифровые антенные решетки (ЦАР) [205], получившие наименование Smart-антенн. В данном контексте слово Smart буквально означает «умные», интеллектуальные антенны. Неслучайно в качестве синонима, соответствующего этому направлению в антенной технике, довольно часто употребляют еще и термин Intelligent Antenna. Возможно, это не самые удачные понятия, однако именно в них как нельзя лучше отражена суть возможностей, предоставляемых технологией цифрового диаграммообразования (digital beamforming), благодаря которым антенны все более приобретают черты систем искусственного интеллекта.

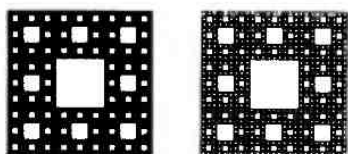


Рис. 9.51. Ковер Серпинского

Следует признать, что на данный момент в отношении антенных систем название Smart пока содержит больше авансов на перспективу, нежели отражает реальное положение вещей. Однако уже в рамках имеющегося научного задела Smart-технологии, реализованные в ЦАР, «умеют» многое.

Каковы же преимущества нового класса антенных систем перед традиционными антеннами, в том числе по сравнению с их прототипом — фазированными антенными решетками, используемыми, например, в Smart-оборудовании связных систем фирмы Paratek (США, www.paratek.com)? Ответ на этот вопрос имеет смысл предварить кратким экскурсом в схемотехнические основы ЦАР [83, 86–88].

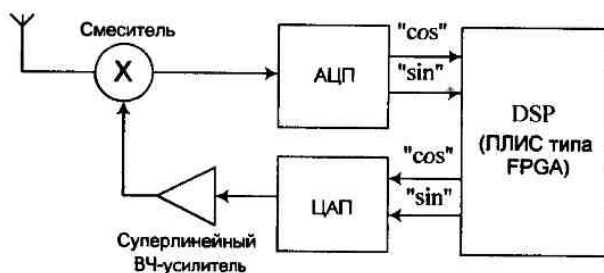
Цифровая антенная решетка — это антенная система, представляющая собой совокупность аналого-цифровых каналов с общим фазовым центром, в которой формирование диаграммы направленности осуществляется в цифровом виде, без использования фазовращателей.

Теоретические основы такого подхода к построению антенн были заложены еще в 60–70-е годы прошлого века, но лишь в настоящее время с развитием микропроцессорной техники начался этап практической реализации накопленного научного задела. Наиболее последовательным приверженцем этого направления явилась киевская научная школа, возглавляемая профессором В.А. Варюхиным.

Данным научным коллективом были созданы теоретические основы многоканального анализа со сверхразрешением сигналов, апробирован ряд макетов и опытных образцов радиотехнических систем с цифровыми антенными решетками. Как следствие, к началу 1990-х годов в СССР существовал уже весьма представительный ряд научных школ, совокупные достижения которых по многим теоретическим и практическим вопросам опережали аналогичные разработки за рубежом.

Современные технологии ЦАР, продвигаемые во всех технически развитых странах мира, по большому счету можно считать венцом развития идеологии программно реконфигурируемой архитектуры в радиотехнике (Software Radio). Суть ее сводится к тщательно согласованной интеграции процессоров цифровой обработки сигналов в виде DSP или программируемых логических интегральных схем FPGA, а также аналого-цифровых и цифроаналоговых преобразователей (АЦП и ЦАП) в рамках одного модуля или даже чипа [86]. Идеализированная схема типовой архитектуры Software Radio представлена на рис. 9.52 [205].

Рис. 9.52. Архитектура Software Radio



Подобное построение каналов ЦАР позволяет избежать аппаратной зависимости процедур обработки сигналов и упрощает их адаптацию под нужды того или иного стандарта функционирования. Как справедливо отмечается в рекламных проспектах фирм-разработчиков, на основе технологии цифрового диаграммообразования может быть получена максимальная унификация узлов и блоков аппаратуры, простота реконфигурации и модификации систем связи, которая сводится зачастую лишь к замене их программного обеспечения. При этом обеспечивается оптимизация и ресурсно-функциональная адаптация архитектуры радиоэлектронного оборудования под конкретно решаемые в текущий момент времени задачи.

В зависимости от области применения и специфики решаемых задач ЦАР могут быть пассивными либо активными. В первом случае имеет место приемная цифровая антенная решетка, во втором — приемопередающая либо разнесенная совокупность приемной и передающей ЦАР. В большинстве прикладных задач предпочтительным является использование разнесенных систем, что вполне может быть выполнено в случае базовых станций сотовой связи. Схематически вариант структуры приемного сегмента ЦАР показан на рис. 9.53.

Структурная декомпозиция приемной ЦАР на составные подсистемы может осуществляться по целому ряду признаков, однако прежде всего в ее составе необ-

ходимо выделить аналоговую и цифровую части, отличающиеся формой представления сигналов. В зависимости от возможностей элементной базы и диапазона функционирования ЦАР аналого-цифровое преобразование сигналов может быть реализовано на промежуточной частоте либо непосредственно на несущей. При этом сигнал такта на все АЦП должен разводиться от общего генератора тактовых импульсов с таким расчетом, чтобы во всех приемных каналах АЦП срабатывали строго синхронно.



Рис. 9.53. Вариант структуры приемной ЦАР

Именно спецификой размещения АЦП схемотехника цифрового диаграммообразования по своей сути заметно отличается от традиционных фазированных решеток (ФАР), в которых АЦП устанавливается лишь на суммарный или разностные каналы приема.

Основной особенностью перспективных схемных решений метрового и соседнего с ним участка дециметрового диапазонов волн является оцифровка сигналов непосредственно на несущей частоте (рис. 9.53). В этом случае АЦП подключаются непосредственно к выходам соответствующих антенных элементов через согласующие усилители. При этом исключаются операции преобразования частоты, детектирования сигналов с выделением их огибающей, что сокращает энергетические потери, повышает чувствительность приемной системы и упрощает аппаратную реализацию.

При оцифровке сигналов на промежуточной частоте в состав аналоговой части каждого из приемных каналов дополнительно вводится смеситель и полосовой усилитель промежуточной частоты, а также в состав ЦАР в целом – общий для всех каналов гетеродин.

Другим ключевым фактором, определяющим схемотехнику аналогового приемного тракта, является способ формирования квадратурных составляющих сигналов. Если расквадратурирование сигналов осуществляется в аналоговом виде, то структура аналогового тракта может строиться по принципу, представленному на рис. 9.54 [84]. Альтернативой такому подходу является лишь цифровое формирование квадратурных составляющих с помощью процедуры дискретной гильбертовской фильтрации разной степени сложности. В простейшем случае для немодулированных по частоте сигналов предусматривается оцифровка через нечетное число периодов заполняющих сигналы радиочастотных колебаний. При этом четные по номеру следования отсчеты фигурируют в последующей обработке как синусная квадратура, а нечетные – в качестве косинусной составляющей.

На окончательный выбор конкретного варианта формирования квадратурных составляющих влияет целый ряд факторов, в том числе принятый к реализации метод дополнительного стробирования отсчетов АЦП. Дело в том, что при высоких частотах дискретизации формирование цифровой диаграммы направленности антенной решетки в темпе поступления отсчетов АЦП зачастую не представляется возможным. Поэтому, как вынужденная мера, используется операция прореживания информационного потока (дополнительного стробирования

цифровых отсчетов). Однако выполняться она должна не путем отбрасывания излишних цифровых данных, а посредством согласованного накопления отсчетов в жестко отведенных интервалах времени — стробах [85]. Именно такая операция является первой процедурой цифровой обработки, которая должна выполняться непосредственно после аналого-цифрового преобразования сигналов.

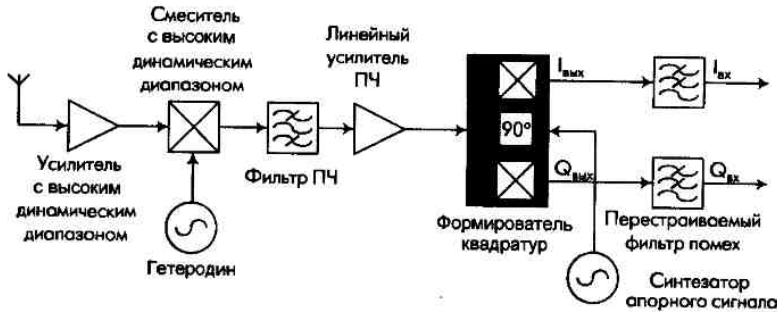


Рис. 9.54. Аналоговая схема формирования квадратур в приемном канале ЦАР

С учетом сказанного в состав цифровой части каждого приемного канала помимо АЦП будет входить устройство дополнительного стробирования отсчетов, именуемое далее цифровым фильтром формирования квадратур (ЦФФК), в силу популярной реализации на его основе дискретной гильбертовской фильтрации.

Реализация ЦФФК возможна в двух вариантах: с одноканальным аналоговым вводом или с двумя квадратурными аналоговыми входами, стыкуемыми с выходами квадратур схемы рис. 9.54. Первый вариант более прост в техническом исполнении, требует установки в каждый приемный канал лишь одного АЦП. Принципиальное же преимущество двухканального ЦФФК состоит в одновременности формирования квадратур, идентичности описания их огибающих при импульсном сигнале, что очень важно при прямоугольной форме импульса. Кроме того, двухканальный вариант имеет в частотной области дифракционный максимум, который отстоит от центральной частоты приема вдвое дальше, по сравнению с одноканальным ЦФФК. Это уменьшает погрешности расквадратирования сигналов и частотные искажения при относительно широкополосных сигналах (для связных задач, например поддержки CDMA, требуется полоса пропускания около 1,23 МГц).

С точки зрения схемотехники цифровой сегмент каждого приемного канала ЦАР представляет собой функционально законченный модуль. Для реализации алгоритма цифровой обработки сигналов в нем предпочтительнее использовать программируемые логические интегральные схемы (ПЛИС), например типа FPGA.

В отличие от сигнальных процессоров (DSP) FPGA позволяют реализовать обработку цифровых данных по схеме жесткого автомата. Это позволяет избежать асинхронизма в работе многоканального измерителя в сочетании с возможностью оперативной замены алгоритмов обработки сигналов посредством перезагрузки требуемой конфигурации FPGA из флэш-ПЗУ. Во многом благодаря указанным достоинствам FPGA-направление в обработке сигналов ныне суще-

ственно потеснило рынок процессорных систем и заняло рекордные по достигнутым показателям позиции.

Для максимального приближения к реализации потенциальных возможностей технологии цифрового формирования лучей характеристики аналоговых сегментов приемных каналов должны обладать высокой идентичностью. Учитывая неизбежность технологических погрешностей, возникающих в процессе изготовления и монтажа приемных модулей, в каждом канале необходимо предусмотреть блок цифровой коррекции его амплитудно- и фазочастотных характеристик под заданный эталон [90]. При достаточно высоком линейном динамическом диапазоне аналогового тракта коррекцию коэффициентов передачи приемных модулей целесообразно производить непосредственно перед формированием цифровой диаграммы направленности. Поэтому указанный блок коррекции в рассматриваемом варианте структуры ЦАР размещается в тех же ПЛИС, что и блок ЦФФК, корректируя квадратурные составляющие результатов дополнительного стробирования отсчетов АЦП (отклики ЦФФК) по загружаемым из управляющего процессорного модуля коэффициентам коррекции. Расчет этих коэффициентов должен проводиться по контрольному (тестовому) сигналу, формирование которого осуществляется генератором контрольного сигнала, общим для всех приемных каналов ЦАР [90]. Подачу тестового воздействия на приемные каналы предпочтительно выполнять на несущей частоте, охватывая контролем максимально возможную часть аналогового тракта.

Полученные в результате операций дополнительного стробирования и коррекции отклики ЦФФК всех приемных каналов далее должны поступать на цифровой формирователь диаграммы направленности антенной решетки. При необходимости информация о мгновенном амплитудно-фазовом распределении электромагнитного поля на раскрые антенной системы в виде отсчетов АЦП либо их частичных сумм предварительно сбрасывается в буферное запоминающее устройство.

Следует отметить, что схемотехника цифровой части приемного сегмента ЦАР существенно зависит от выбора интерфейса межмодульной трансляции данных, являющегося самостоятельным структурным элементом. С точки зрения удобства совместимости с программным обеспечением компьютеров линии Intel PC предпочтительными являются решения в классе интерфейсов PCI и CompactPCI (сPCI). Впрочем в приложениях, критичных к условиям эксплуатации, как правило, PCI-концепция не получает развития, и выбор должен происходить в пользу более дорогих сPCI-модулей.

Цифровое диаграммообразование осуществляется программным способом с помощью цифровых сигнальных процессоров (DSP) либо программируемых матриц логических элементов (FPGA), выполняющих синфазное взвешенное суммирование напряжений всех первичных приемных каналов для интересующих угловых направлений.

Особенностью схемотехники приемопередающих ЦАР, используемых в связанных приложениях, является наличие наряду с приемным сегментом аналогичной по составу передающей подсистемы, а также общей для них антенной решетки.

Цифровой передающий сегмент активной ЦАР идентичен приемному с точностью до замены в нем АЦП на ЦАП и характеризуется инверсией алгоритмов об-

работки сигналов. Кроме того, существуют отличия и в реализации аналогового сегмента передающей системы, связанные с необходимостью усиления сигналов по мощности, переноса их по частоте и фильтрации побочных излучений.

При цифровом формировании зондирующего луча синтезированный с помощью DSP или FPGA массив цифровых напряжений зондирующего сигнала поступает на ЦАП, с выходов которых после усиления мощности, фильтрации и переноса по частоте аналоговые сигналы подаются на запитку соответствующих антенных излучателей.

Рассмотренный принцип цифрового диаграммообразования является достаточно общим, и в конкретных устройствах может претерпевать определенные вариации, обусловленные возможностями элементной базы, конкретным опытом разработчиков и их теоретической подготовкой, объемами финансирования того или иного проекта.

Ключевой особенностью функционирования ЦАР является, как уже отмечалось, цифровое формирование лучей (ЦФЛ) характеристики направленности (ХН), которое в задачах связи позволяет эффективно реализовать динамическую оптимизацию обслуживаемой зоны покрытия на основе оперативного перенацеливания цифровых приемных лучей по группам корреспондентов (рис. 9.55) в зависимости от территориального распределения абонентов сотовой связи. Созвездие приемных лучей, синтезируемое, например, по алгоритмам быстрого преобразования Фурье либо посредством классических процедур дискретного Фурье-анализа, является, по сути, совокупностью «пространственно-частотных фильтров», каждый из которых селективирует строго определенный набор сигналов и подавляет остальные, одновременно принимаемые в широком пространственном секторе, как помеховые.

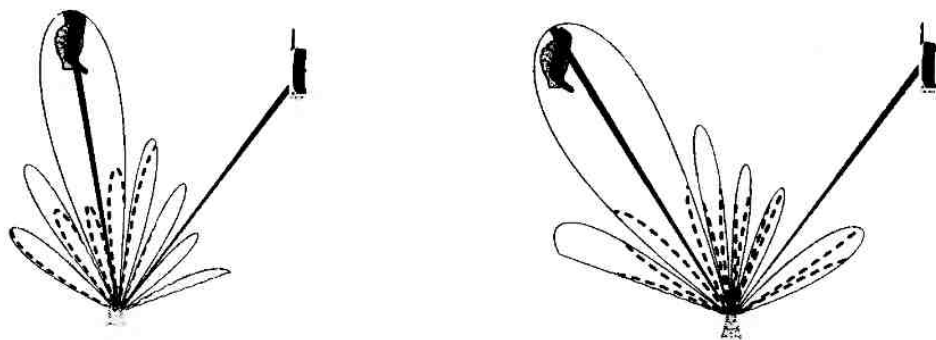


Рис. 9.55. ЦАР позволяет адаптивно перенацеливать лучи на мобильных абонентов

При этом существенно улучшается качество связи в условиях многолучевого распространения радиоволн, а также резко повышается помехозащищенность системы в случае интенсивного радиопротиводействия. Предпосылкой тому служит цифровое формирование в программно-реконфигурируемых каналах высокочастотных фильтров на выходах приемных устройств в совокупности с наращиванием динамического диапазона при когерентном накоплении

отсчетов напряжений в процессе пространственно-временной обработки сигналов. В результате, формирование нулей ХН в направлениях источников помех (рис. 9.56) осуществляется без «заплывания» провалов, обычно возникающего при недостаточном динамическом диапазоне приемного модуля. Многочисленные эксперименты подтверждают возможность подавления активной шумовой помехи в восьмиэлементной ЦАР не только по боковым лепесткам, но и в главном луче ХН более чем на 30 дБ при среднеквадратическом отклонении коэффициентов усиления аналоговых приемных каналов 0,5 дБ и величине фазовых ошибок не более 3° [88].

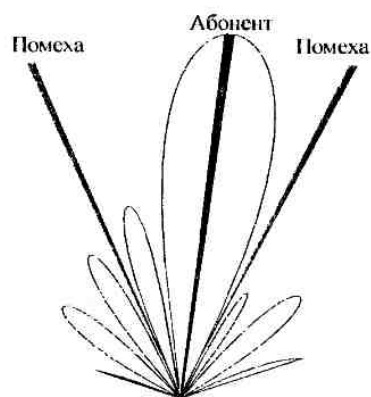


Рис. 9.56. Адаптивное формирование нулей ДН ЦАР в направлениях помех

Немаловажным достоинством использования ЦАР на базовых станциях является также увеличение пропускной способности сотовой связи за счет одновременного многолучевого приема сигналов во всем рабочем секторе. Реализация сверхрэлееской разрешающей способности по направлениям прихода сигналов, частоте и времени задержки способствует в этом случае достижению требуемой детализации панорамных срезов пространственной обстановки.

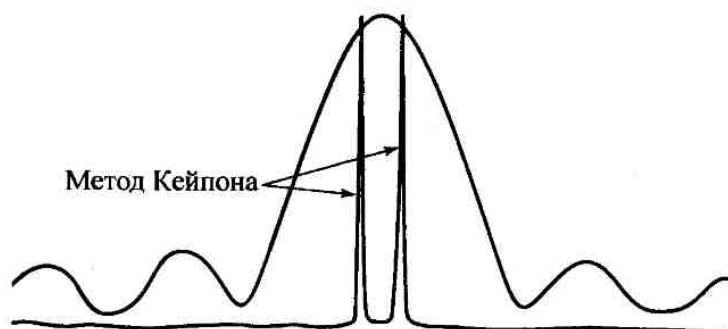
В качестве примера на рис. 9.57 проиллюстрирован эффект сверхрэлееского разрешения по угловой координате двух точечных источников в линейной ЦАР с помощью алгоритма Кейпона [60].

Нетрудно заметить, что традиционная обработка (верхняя кривая рис. 9.57) не позволяет получить визуальное разрешение, тогда как процедура Кейпона отчетливо выдает два обособленных отклика. Следует подчеркнуть, что в настоящее время известно достаточное количество методов сверхразрешения, реализация которых в зависимости от отношения сигнал/шум позволяет разделять до десяти и более точечных объектов в пределах парциального лепестка приемной диаграммы направленности ЦАР. Однако для решения непосредственно связанных задач, исключая процесс вхождения в связь, предпочтительным является все же метод максимального правдоподобия (МП), из которого все остальные следуют как квазиоптимальные варианты. Именно при использовании метода МП цифровое формирование луча обеспечивает близкие к потенциальным точность оценивания параметров сигналов и их разрешение.

Перечисленные, а также ряд других преимуществ ЦАР послужили причиной их стремительно растущей рыночной привлекательности. Пройдя в конце 90-х годов этап демонстрационных проектов [87], сегодня данная технология вступила в стадию серийного выпуска антенных систем для существующей и перспективной инфраструктуры сотовой связи.

Среди фирм, первыми разработавших и освоивших производство ЦАР как серийной продукции, ведущие позиции сегодня занимают ArrayComm Inc., Metawave Communications Corp., AirNet Communications Corp., Wireless Online Inc. (США), Ericsson.

Рис. 9.57. Эффект
сверхрешеевского
разрешения двух
сигналов по методу
Кейпона



Первенство в освоении серийного производства ЦАР для базовых станций стандарта CDMA принадлежит американской компании Metawave Communications Corp., представленной на рынке интегрированных Smart-антенн их семейством с торговой маркой Spotlight. Первые ЦАР от Metawave Communications, получившие наименование Spotlight 2000 (2100), поддерживали только один диапазон несущих 800–900 МГц. Затем было применено двухдиапазонное решение за счет добавления полосы частот 1800–1900 МГц (Spotlight 2200).

Типовая Smart-антенна базовой станции от Metawave состоит из 12-элементного массива излучателей, устанавливаемых на базовой станции обычно по трехсекторной схеме (рис. 9.58).

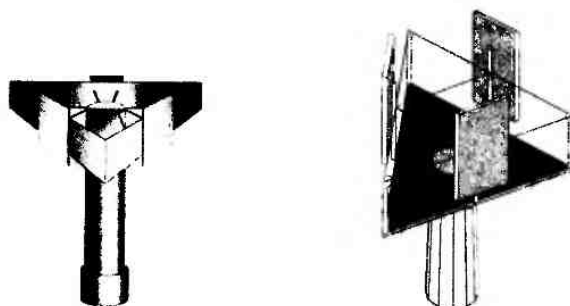


Рис. 9.58. Типовая Smart-антенна базовой станции от Metawave

Каждая секторная решетка сформирована из четырех антенных элементов (рис. 9.59), пример ориентации характеристик которых представлен на рис. 9.60.

Ширина характеристики направленности каждого антенного элемента на уровне -3 дБ составляет около 30° . Благодаря процедуре ЦФЛ за счет управления фазой и мощностью сигнала в антенных каналах результирующая ширина парциальных лучей каждого секторного сегмента ЦАР может принимать значения 180, 120 или 60° (рис. 9.61). При этом сам сектор излучения (приема) может смещаться влево-вправо относительно физической нормали на угол $\pm 30^\circ$, а также адаптивно изменяться геометрическая форма самой соты, приобретая трех-, четырех- и даже шестисегментные очертания (рис. 9.61).

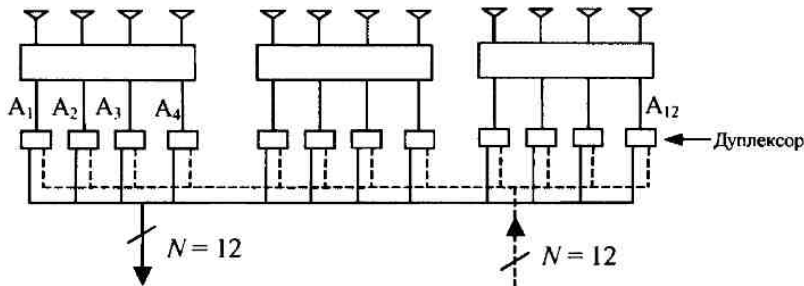


Рис. 9.59. Трехсекторная антенная решетка семейства Spotlight

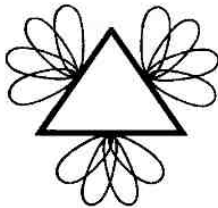


Рис. 9.60. Ориентация лучей ДН антенны Spotlight

Адаптивная обработка сигналов антенны Spotlight позволяет в каждом из 120-градусных секторов вырезать до трех подсекторов, пораженных помехами, в реальном масштабе времени, осуществляя так называемый динамический синтез сектора (DSS). Все это позволяет оператору базовой станции рационально перераспределять часть незадействованного трафика в те области, где в таковом возникает острая потребность, и упрощает частотное планирование сети. Фактически конфигурацию лучей ХН в каждом из рабочих секторов можно независимо оптимизировать под конкретную сотовую нагрузку (рис. 9.62) в соответствии со спецификой окружающего ландшафта, статистикой распределения запросов по времени суток, дням недели, в случае непредвиденных форс-мажорных обстоятельств.

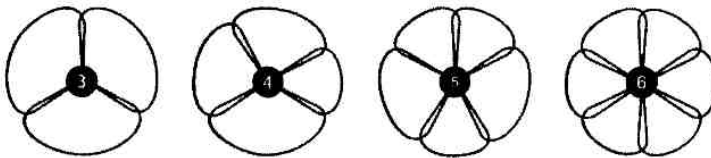


Рис. 9.61. Адаптивное изменение количества и размеров рабочих секторов в ЦАР

Увеличение количества виртуальных секторов приема с трех до шести позволяет существенно поднять емкость сотовой ячейки (количество обслуживаемых абонентов). По данным Metawave [226], такой прирост в стандарте CDMA 2000 на основе Spotlight-решений может достигать 94%.

Аппаратная реализация Spotlight-концепции базируется на использовании заказных ИС (ASIC) и FPGA, что лишний раз подчеркивает справедливость вывода о предпочтительности применения FPGA по сравнению с DSP при решении задач цифрового формирования луча [86].

Одна из последних разработок Metawave — антенный комплекс Spotlight 2230 — представляет собой аппаратуру ЦАР, интегрируемую в состав базовой станции CDMA 1X совместно с оборудованием базовой станции фирмы Lucent Flexent

Modular Cell Base Station [226]. При этом фирма Lucent поставляет ультранинейные усилители мощности для передающего сегмента и малошумящие усилительные модули для многоканального приемника. На рис. 9.63 схематически приведен внешний вид типового оборудования базовой станции CDMA с ЦАР Spotlight 2230.

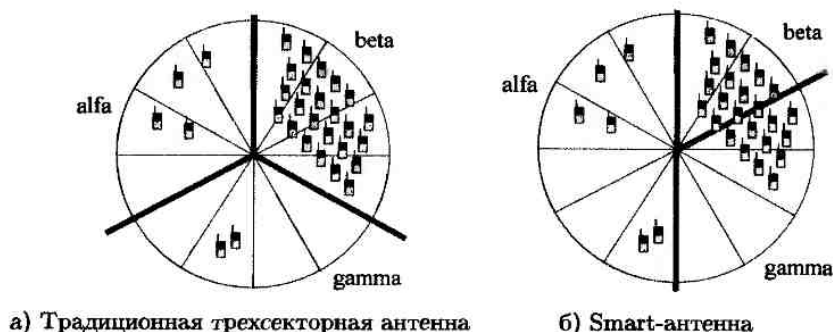


Рис. 9.62. Динамический синтез секторов в зависимости от распределения сотовой нагрузки

Специфика цифрового оборудования ЦАР позволяет интегрировать ее в базовую станцию любого стандарта с минимумом стыковочных работ. При этом стоимость дополнительного оборудования, развертываемого на базовой станции, по данным [226], позиционируется на уровне 10% от стоимости аппаратуры типовой базовой станции CDMA, рассчитанной на обслуживание 48–72 одновременных вызовов. Однако, учитывая, что применение ЦАР существенно повышает емкость сети, речь может идти о значительной экономии средств за счет отказа от необходимости развертывания дополнительных базовых станций для достижения той же емкости.

В отличие от компании Metawave Communications другая американская компания — AirNet Communications (www.aircom.com) — сосредоточила свои усилия на разработке Smart-антенн для модернизации базовых станций стандартов GSM (900, 1800, 1900 МГц) с выходом на поддержку в них спецификаций GPRS, EDGE и даже W-CDMA. В активе AirNet имеется около 70 действующих патентов, обстоятельно страхующих ее интересы в развитии столь нового технологического направления, которые она получила с момента появления на рынке телекоммуникаций в 1994 году.



Рис. 9.63. Типовое оборудование базовой станции CDMA с ЦАР

Технология цифрового диаграммообразования в адаптивных антенных решетках в продукции AirNet представлена базовой станцией, зарегистрированной под торговой маркой AdaptaCell Super Capacity, переводимой дословно как «суперъемкая адаптивная сота».

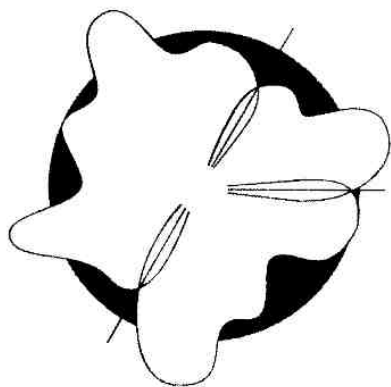


Рис. 9.64. «Интеллектуальная» сота

Опираясь на рекламные материалы, которые читатели могут заказать на сайте фирмы, несложно предположить, что указанная базовая станция обслуживает 8-элементную ЦАР, которая работает в секторе 120° . В таком варианте компоновки полноценная трехсекторная антенная решетка будет содержать 24 элемента. AdaptaCell обеспечивает цифровое представление сигналов с полосой до 40 МГц, что в 35 раз превышает объем данных традиционных базовых станций [203]. Чувствительность приемника составляет -110 дБ, полоса пропускания 200 кГц. Излучаемая мощность заявлена на уровне 20 Вт. Предусмотрена возможность поддержки до 12 несущих. Комплект оборудования базовой станции состоит из входного (Indoor BTS) и выходного (Outdoor BTS) трансиверов. Габариты Indoor-модуля $2,08 \times 0,56 \times 0,81$ м, а Outdoor — $1,73 \times 1,83 \times 1,04$ м, потребляемая мощность соответственно составляет 2,51 и 2,81 кВт.

Примечательно, что базовая станция AdaptaCell Super Capacity совместима с технологией «интеллектуальной соты» (IntelliCell) [122], продвигаемой на рынке Smart-антенн компанией ArrayCom (Сан-Хосе, Калифорния). На рис. 9.64 проиллюстрирован вариант адаптации конфигурации сотовой ячейки под конкретные условия размещения и работы базовой станции. Вместе с Metawave Communications ArrayCom является пионером в разработке Smart-антенн для радиосвязи. Однако, в отличие от Metawave, ArrayCom изначально ориентировалась на рынок 2,5G и 3G систем, поддерживающих стандарт W-CDMA. По этой причине вскоре интересы AirNet и ArrayCom стали пересекаться, и они вынуждены были зачастую объединять усилия, прибегая в ряде случаев к совместной поставке оборудования.

На этом примере отчетливо проявилась тенденция превращения рынка Smart-антенн в арену острой конкурентной борьбы, где ключевые игроки неизбежно вторгаются в круг интересов друг друга. Тем не менее даже столь жесткий бизнес несомненно стоит свеч, ибо, согласно заявлениям разработчиков ArrayCom, интеллектуальная сота IntelliCell на основе технологии цифрового формирования луча позволяет почти вдвое сократить необходимое для покрытия района обслуживания количество базовых станций, на четверть снизить расходы, требуемые для развертывания вновь создаваемой сетевой инфраструктуры, и вдвое — время на выполнение инсталляционных работ. Если верить рекламным документам, то к настоящему времени ArrayCom уже поставила во всем мире около 60 тыс. базовых станций, в том числе для японской системы персональных радиотелефонов (PHS). Планируемый срок окупаемости затрат на внедрение Smart-антенн

составляет год и менее. Следует при этом учесть, что технология ЦАР пока еще находится лишь в начале своего становления. Еще более внушительные результаты по увеличению канальной емкости и размеров покрываемой зоны могут быть достигнуты по мере ее совершенствования, например путем увеличения количества антенных элементов в одном секторе ЦАР до 16 в азимутальной плоскости и до 4–8 — в угломестной.

9.7.2 Методы пространственной селекции сигналов и подавления помех в ЦАР

Основной причиной неоспоримых преимуществ ЦАР перед ФАР в вопросах помехозащищенности и полноты реализации потенциальных возможностей цифровой обработки сигналов является недостаточная точность фиксации амплитудно-фазового распределения электромагнитного поля по раскрыву ФАР, осуществляемая с помощью фазовращателей (ФВ), разрядность которых обычно не превышает 6–7 бит. В ЦАР же для этих целей могут использоваться АЦП и ЦАП с вдвое большей разрядностью (до 16 бит), что позволяет, в частности, формировать более глубокие провалы в диаграмме направленности ЦАР в направлениях источников помех, точнее измерять угловые координаты абонентов на этапе вхождения в связь. Другой аспект состоит в большей неидентичности ФВ по сравнению с характеристиками АЦП и ЦАП, что также приводит в ФАР к дополнительным энергетическим потерям.

Обработка принятых сигналов в ЦАР систем связи делится на два этапа: вхождение в связь и прием данных. В каждом из указанных режимов должны применяться свои алгоритмы обработки. В режиме вхождения в связь в начале анализируется помеховая обстановка путем пеленгации источников активных помех, если таковые имеются, и определяются весовые коэффициенты для последующего вычитания помеховых откликов из напряжений вторичных каналов.

Пеленгация помех выполняется методами спектрального оценивания:

- а) оптимальным (максимум правдоподобия):

$$L = \text{tr}[G \cdot \tilde{R}] \rightarrow \max,$$

где $G = P[P^*P]^{-1}P^*$, $\tilde{R} = UU^*$, P — матрица неизвестных значений ДН ЦАР в направлениях источников помех, U — вектор напряжений сигнальной смеси по выходам приемных каналов ЦАР,

или с учетом свойства векторов $\text{tr}[UU^*] = U^*U$:

$$L = \text{tr}[G \cdot \tilde{R}] = \text{tr}[GUU^*] = U^* \cdot G \cdot U \rightarrow \max \quad (9.1)$$

(итерационный поиск максимума выражения (9.1) изменением матрицы G на множестве оценок параметров сигналов требует меньше вычислительных затрат),

- б) квазиоптимальными (MUSIC, Кейпона, ESPRIT и др.),
в) классическими (метод максимума диаграммы направленности, моноимпульсный метод и др., основанные на электронном сканировании диаграммой направленности в пространстве).

Расчет весовых коэффициентов для формирования «нулей» диаграммы направленности ЦАР в направлениях прихода помех осуществляется с помощью известного алгоритма, матричная запись которого может быть представлена в виде:

$$\tilde{G} = 1 - P \cdot (P^* \cdot P)^{-1} \cdot P^*, \quad (9.2)$$

где 1 — единичная матрица (на главной диагонали единицы, а остальные элементы — нулевые), а в качестве матрицы P может использоваться матрица, столбцы которой образованы векторами волнового фронта помеховых сигналов с аргументами комплексных экспонент, соответствующими значениям угловых координат, вдоль которых должны ориентироваться «нули» передаточной функции ЦАР; * — символ операции комплексно-сопряженного транспонирования матрицы.

Например, в случае одного помехопостановщика, действующего с углового направления θ град. относительно нормали к четырехэлементной решетке, матрица P имеет вид:

$$P = \begin{bmatrix} 1 \\ e^{j \cdot x} \\ e^{j \cdot 2 \cdot x} \\ e^{j \cdot 3 \cdot x} \end{bmatrix}, \quad \text{где } x = \frac{2\pi d}{\lambda} \sin(\theta \frac{\pi}{180}).$$

При этом весовой множитель для формирования провала в диаграмме направленности рассчитывается согласно выражению:

$$\tilde{G}_1 = (1 - \frac{1}{R} P \cdot P^*),$$

где R — количество антенных элементов в решетке, в данном случае равное 4.

При воздействии двух помехопостановщиков с известными координатами θ_1 , θ_2 , полученными в режиме анализа помеховой обстановки, матрица P запишется в виде:

$$P = \begin{bmatrix} 1 & 1 \\ e_1^{j \cdot x} & e_2^{j \cdot x} \\ e^{j \cdot 2 \cdot x_1} & e_2^{j \cdot 2 \cdot x_1} \\ e^{j \cdot 3 \cdot x_1} & e^{j \cdot 3 \cdot x_2} \end{bmatrix}, \quad \text{где } x_1 = \frac{2\pi d}{\lambda} \sin(\theta_1 \frac{\pi}{180}), \quad x_2 = \frac{2\pi d}{\lambda} \sin(\theta_2 \frac{\pi}{180}).$$

В случае произвольного количества каналов ЦАР задача формирования матрицы P для расчета весовых коэффициентов подавления помех обобщается добавлением строк снизу, а при увеличении количества помехопостановщиков — добавлением столбцов справа.

Вычитание помеховых сигналов из напряжений приемных каналов производится на основе информации об угловом положении помехопостановщиков согласно выражению:

$$U_b = \tilde{G} \cdot U_0, \quad (9.3)$$

где U_0 — исходный вектор напряжений, U_b — он же после вычитания откликов помех.

На завершающем этапе процесса вхождения в связь по известным угловым координатам абонентов осуществляется ориентация максимумов ДН ЦАР в сторо-

ну ожидаемого направления приема полезных сигналов. При этом в направлении помехи должен по-прежнему формироваться провал диаграммы направленности. С этой целью вектор напряжений, полученный после процедуры вычитания помех, фазуруется с помощью сигнального вектора S согласно выражению:

$$U_b = S^* \cdot \tilde{G} \cdot U_0 = S^* (1 - P \cdot (P^* \cdot P)^{-1} \cdot P^*) U_0, \quad (9.4)$$

где $S = \begin{bmatrix} 1 \\ e^{j \cdot \gamma} \\ e^{j \cdot 2 \cdot \gamma} \\ e^{j \cdot 3 \cdot \gamma} \end{bmatrix}$ — сигнальный вектор, $\gamma = \frac{2\pi d}{\lambda} \sin(\beta \cdot \frac{\pi}{180})$, β — угловая координата (в градусах) абонента, в направлении которого необходимо сформировать максимум диаграммы направленности антенной решетки при вхождении в связь.

Для демонстрации работоспособности процедуры подавления помех (9.3) и последующего формирования максимума ДН в требуемом направлении (9.4) можно привести результаты математического моделирования в пакете Mathcad, представленные на рис. 9.65, 9.66.

При этом угловая координата полезного сигнала была задана равной -20° , а одиночной помехи $+20^\circ$ относительно нормали к четырехэлементной решетке. Амплитуда сигнала составляла 10 квантов АЦП, а помехи — 1000, разность фаз сигнала и помехи равнялась 45° . Нетрудно заметить, что в результате вычитания помехи происходит искажение сигнального отклика, которое растет по мере сближения угловых координат провала диаграммы направленности и направления на полезный сигнал (см. рис. 9.67).

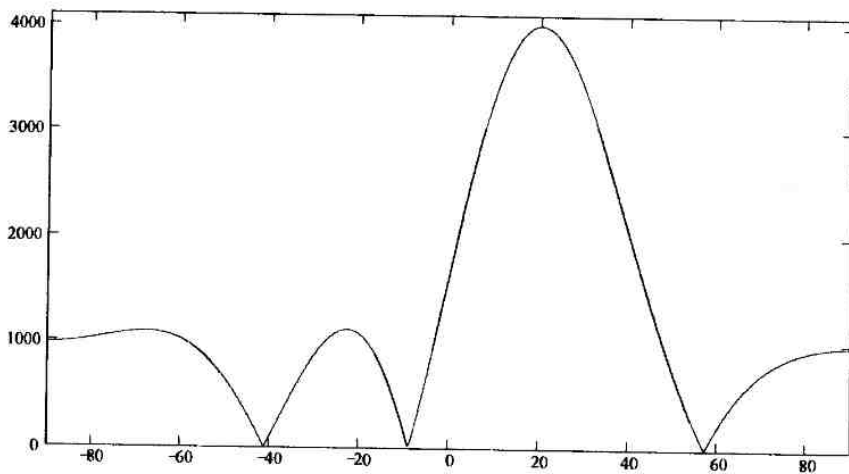


Рис. 9.65. Сигнальная смесь до подавления помехи

Следует отметить, что глубина подавления помехи определяется не только разрядностью фазирующих коэффициентов, но и точностью измерения координат помехопостановщика. При этом погрешность в десятые доли градуса уже приводит к недопустимым искажениям диаграммы направленности после подав-

ления сигнала, вызванным нескомпенсированными остатками помехи. В качестве примера на рис. 9.68 в виде верхней кривой показан отклик четырехэлементной решетки после вычитания одиночной помехи, погрешность измерения координат источника которой составила $0,5^\circ$. Для снижения влияния ошибок пеленгации на качество подавления помехи может быть рекомендовано использование двухстолбцевой матрицы P при формировании весового вектора, при этом в качестве аргументов экспонент должны задаваться угловые координаты левой и правой границ сектора возможного нахождения помехи (например, измеренная угловая координата за вычетом СКО ошибки измерения и с добавлением этого СКО). Об эффективности такого подхода позволяет судить штриховая линия на рис. 9.68, описывающая отклик решетки при формировании двойного провала в направлении помехи (20°). При этом координаты провалов симметрично смещены относительно пеленга помехи на величину $\pm 1^\circ$. Точечной линией по-прежнему представлен беспомеховый сигнальный отклик.

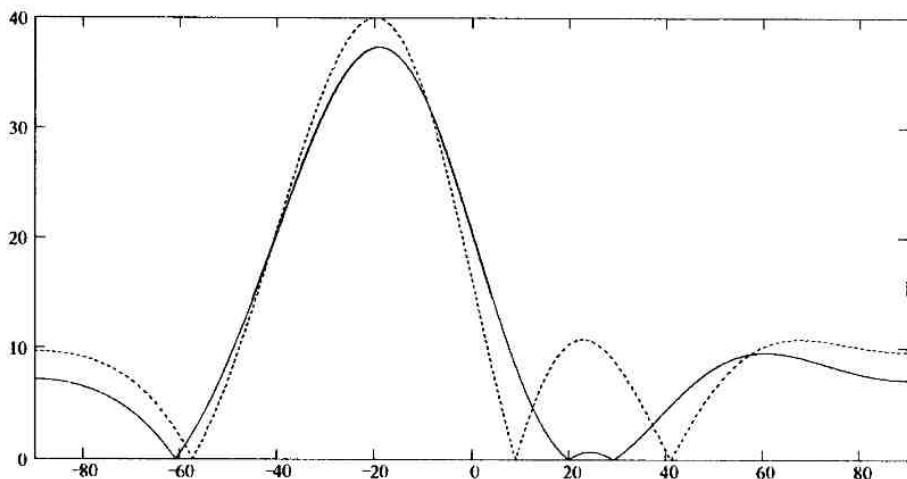


Рис. 9.66. Сигнальная смесь после подавления помехи (сплошная линия) и отклик решетки в беспомеховой ситуации (пунктир)

В целом результаты многих демонстрационных проектов свидетельствуют о возможности достижения подавления широкополосной помехи рассмотренным способом на величину не менее 30–40 дБ.

При факторизуемых диаграммах направленности сказанное справедливо и для раздельного функционирования помехозащитных процедур в ортогональных угловых плоскостях.

9.8 ММО-системы на базе ЦАР

В зависимости от наличия ЦАР в канале связи и того, в роли приемной или передающей антенны она используется, различаются следующие их типы:

SISO (Single Input-Single Output) — «одиночный вход — одиночный выход», на передающем и приемном терминалах используются обычные антенны;

SIMO (Single Input-Multiple Output) — «одиночный вход — множественный выход», в канале связи имеется передающая ЦАР;

MISO (Multiple Output-Single Input) — «множественный вход — одиночный выход», используется приемная ЦАР;

MIMO (Multiple Input-Multiple Output) — «множественный вход — множественный выход», ЦАР находится на приемном и передающем терминалах.

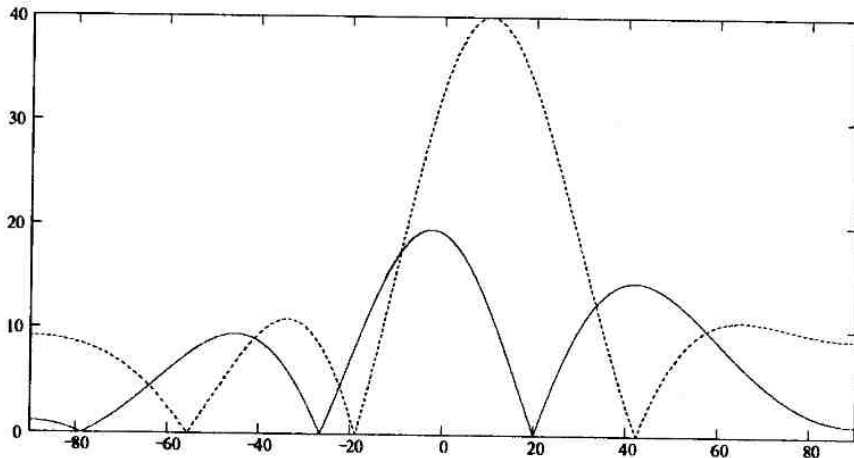


Рис. 9.67. Искажение сигнала (сплошная линия) после вычитания помехи при разnose их угловых координат 10° (источник помехи находится на 20°), пунктирная линия — сигнал в отсутствие помех

В целом такая классификация весьма непрактична и рассматривается здесь лишь как дань сложившейся за рубежом традиции. Дело в том, что в зависимости от режима функционирования ЦАР ее можно рассматривать и как одиночный излучатель, и как множественный. Неслучайно с развитием технологии Smart-антенн в последнее время в рассмотренную классификацию стали вкладывать несколько иной смысл, распространяя ее исключительно на каналы связи, содержащие ЦАР на обоих конечных пунктах, причем эта классификация касается лишь режимов функционирования таких антенн. Так, в случае использования Smart-антенны для формирования единственного луча на передачу и на прием (сопровождение одиночных пользователей антенной базовой станции или канал связи «точка-точка»), когда все каналы участвуют в формировании единственного излучаемого сигнала и аналогично принимают его с одного направления, речь идет об эквиваленте системы SISO.

Если прием может осуществляться одновременно с нескольких направлений по множеству синтезированных лучей, как, например, на базовой станции, а передатчик излучает один сигнал, то фактически имеет место режим MISO. Этот вариант также охватывает случай приема одиночного сигнала в условиях помех,

возникающих за счет многолучевого распространения радиоволн, когда в направлениях помех необходимо сформировать нули диаграммы направленности.

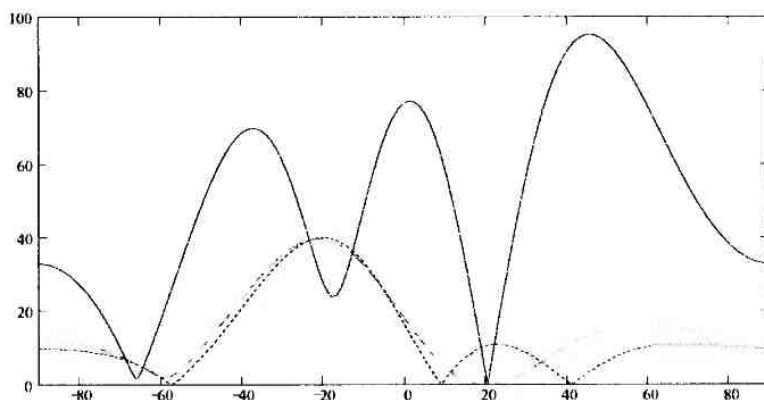


Рис. 9.68. Отклик четырехэлементной решетки после вычитания одиночной помехи (верхняя кривая), отклик решетки при формировании двойного провала в направлении помехи (20°) (штриховая линия), беспомеховый сигнальный отклик (точечная линия)

При излучении сигналов сразу в нескольких направлениях для связи с абонентскими терминалами, формирующими одиночные лучи в направлении базовой станции, передающая ЦАР функционирует в режиме SIMO.

Наконец, режим MIMO соответствует работе ЦАР с формированием множества лучей на прием и на передачу.

Отличительной особенностью рассмотренных выше вариантов применения передающих ЦАР является то, что все их каналы участвуют в формировании излучаемых сигналов. Однако возможен также вариант, когда каждый парциальный канал передающей ЦАР при их совместном излучении функционирует в качестве отдельного, независимого источника сообщения [129], излучая свой автономный сигнал без электромагнитной кооперации с остальными каналами (рис. 9.69). При этом поток данных на передающей стороне перед излучением в эфир разделяется на независимые последовательности битов, которые далее пересылаются на приемный пункт одновременно, используя структурную параллельность массива антенн. Соответствующая технология получила наименование пространственного мультиплексирования (иногда в отношении ее употребляется также термин пространственно-временного кодирования, Space-Time Coding).

В последние годы именно данный режим находится в центре внимания разработчиков и даже планируется в качестве основы для будущего стандарта IEEE 802.11n (а также описан в стандарте IEEE 802.16-2004). Именно такой вариант функционирования приемной и передающей ЦАР стал однозначно ассоциироваться большинством специалистов с системой MIMO. Поэтому далее по тексту под MIMO будем понимать совокупность передающей ЦАР, работающей в режиме автономного излучения сигналов по каждому передающему каналу, и приемной ЦАР с совместной (кооперативной) обработкой всех сигналов по выходам ее приемников.

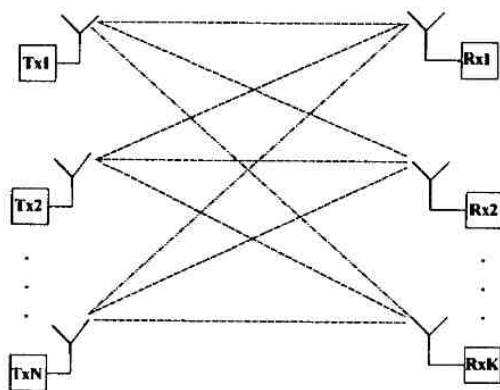


Рис. 9.69. Система ММО

имное влияние элементов и снизить кросс-поляризационные помехи.

Для ММО-антенн в силу их многоканальности вопрос минимизации габаритов излучателей стоит довольно остро. Поэтому в [196] предложено использовать симметричные печатные инверсные F-антенны (IFA) с дополнительными емкостными и индуктивными нагрузками. О роли емкостных нагрузок уже говорилось в подразделе 8.3.4, использование же индуктивной нагрузки в виде меандра позволяет на 30% сократить размеры IFA.

Конструктивное исполнение антенных элементов типовых абонентских ММО-систем базируется на рассмотренных в предыдущих пунктах данной главы технических решениях. Например, это могут быть четыре вибраторные антенны, размещенные на обратной стенке дисплея ноутбука.

Другой вариант предполагает использование четырех РІFA-антенн с Е-образными вырезами в горизонтальной пластине [225]. Каждая Е-пластина окружена заземленными вертикальными секциями, что позволяет уменьшить вза-



Рис. 9.70. Двухдиапазонная ММО-антенна

Как известно, в настоящее время в рабочих группах по разработке стандарта 802.11n идет дискуссия по поводу выбора в качестве ММО-антенны двухэлементной либо четырехэлементной цифровой антенной решетки. Рисунок 1.70 может служить прототипом компромиссного решения, в котором объединены оба альтернативных варианта, правда для работы в разных диапазонах (в данном случае двухэлементная решетка обслуживает канал 900 МГц, а четырехэлементная — канал 1800 МГц) [225].

Кстати, данные решетки являются двухполяризационными, что позволяет повышать эффективность работы средств связи в условиях многолучевого распространения радиоволн, проиллюстрированных на рис. 9.71, а при независимой передаче сигналов по каждой из поляризационных составляющих — дополнительно удвоить пропускную способность канала связи. При этом в антенной решетке рис. 9.70 достигнуто подавление кроссполяризационной помехи на 34–39 дБ.

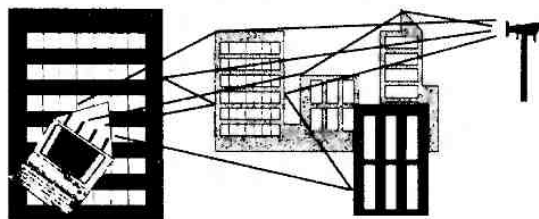


Рис. 9.71. Многолучевое распространение радиоволн в системе ММО

решения источников требуется столь значительное отношение сигнал-шум, что дальность связи резко сокращается. Двухэлементный же вариант, в том числе полученный из четырехканальной ЦАР попарным объединением ее элементов, создает более простые условия для разрешения сигналов. Он позволяет заметно увеличить дальность связи и даже выйти на ту же пропускную способность, что и в четырехэлементной ЦАР, за счет меньших ошибок измерения квадратурных составляющих амплитуд сигналов и, как следствие, возможности использования более высокоразрядной QAM-модуляции.

Впрочем, решающим фактором в окончательном выборе из альтернативы «2–4» может стать не дальность связи и не пропускная способность канала, а его помехозащищенность. С этой точки зрения четырехэлементная приемная ЦАР имеет бесспорные преимущества, поскольку позволяет сформировать при работе по схеме «2 передатчика–4 приемника» два нуля в ДН приемной ЦАР в направлениях на источники помех (переотражений).

При этом двухканальная передача на основе четырехэлементной ЦАР становится более адаптивной по сравнению с четырехканальным излучением. К примеру, для обеспечения максимальной дальности связи на излучение могут работать только два внешних элемента, пространственный разнос которых максимальный. Промежуточным вариантом является уже упомянутое попарное объединение элементов.

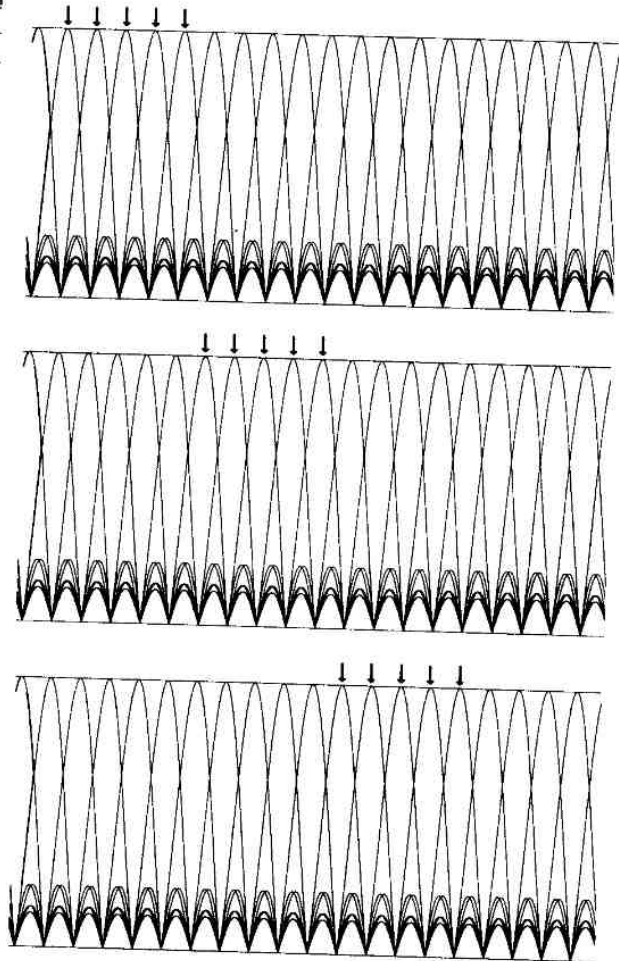
Поскольку расстояние между излучателями в антенных решетках, как правило, не превышает нескольких длин волн несущей частоты, используемой для передачи, при решении задачи измерения параметров сигналов, одновременно формируемых каждым антенным элементом передающей ЦАР, угловое разрешение источников в достаточно большом диапазоне дальностей в рэлеевском смысле неприменимо. Для их же сверхрэлеевского разрешения по угловой координате при таком разноразносе источников требуется слишком большое отношение сигнал-шум. Именно по этой причине дальность действия ММО-систем с одно-

Острая полемика вокруг выбора двух- или четырехэлементных ЦАР для ММО-антенн имеет под собой достаточно веские основания. С одной стороны, четырехэлементная ЦАР — это потенциальная возможность учетверить пропускную способность. Однако при отсутствии ортогональности сигналов по частоте и времени прихода для работы процедур сверхрэлеевского раз-

временным независимым излучением сигналов по парциальным каналам является крайне незначительной и ограничивается обычно сотней метров.

Чтобы повысить точность оценивания квадратурных составляющих амплитуд, в [89] предложено использовать в передающих каналах ортогональные по частоте сигналы по аналогии с концепцией метода OFDM (рис. 9.72). При этом для всех передающих модулей задается одинаковой длительность фрейма OFDM, а допустимое количество сигналов T в каждом из них определяется отношением количества точек K быстрого преобразования Фурье (БПФ), используемого для формирования в приемной ЦАР сетки частотных фильтров, к количеству задействованных пространственных каналов передачи N , т. е. $T = K/N$.

Рис. 9.72. Ортогональное расположение частот сигналов в передающих каналах ЦАР



Существенно, что для такого метода связи номиналы частот сигналов в каналах передающей ЦАР не должны повторяться (рис. 9.72). Это позволяет раздельно измерять в приемной ЦАР квадратурные составляющие амплитуд сигналов, сформированных парциальными излучателями активной решетки, при

сколь угодно сильном сближении антенных элементов передающей ЦАР по обобщенной угловой координате. В результате приемные и передающие ЦАР, образующие ММО-канал, можно весьма значительно разносить в пространстве. Данный разнос ограничен лишь требуемой точностью оценки квадратурных составляющих (отношение сигнал-шум).

В [89] приведены результаты моделирования ортогональной по частоте передачи сигналов в ММО-системе. В пакете Mathcad моделировалась работа трехэлементной активной ЦАР, причем по каждому из передающих каналов имитировалось излучение восьми ортогональных по частоте сигналов (рис. 9.73). Рисунок 9.73, а иллюстрирует распределение трех групп сигналов вдоль частотной оси по выходам 64 БПФ-фильтров, синтезированных в каждом вторичном пространственном канале приемной ЦАР. На рис. 9.73, б изображен пространственный отклик приемной ЦАР, представлявшей собой линейку из восьми антенных элементов, на сигналы трех излучающих каналов передающего терминала, причем по горизонтальной оси отложены значения угловой координаты в градусах относительно нормали к приемной решетке. Поскольку угловые расстояния между антенными элементами составляют половину длины волны, то, как и следовало ожидать, результирующий отклик на них приемной антенны не содержит провалов, т. е. разрешение по направлению прихода в данном случае возможно только на основе сверхрэлеевского разрешения. Однако ситуация облегчается тем, что сигналы могут быть разрешены по частоте. На рис. 9.73, в показан трехмерный вид результирующего отклика приемной ЦАР в координатах «угол-частота». При увеличении количества каналов в передающей ЦАР картина по угловой координате не претерпит существенных изменений по сравнению с рис. 9.73, б, тогда как в частотной области множество ортогональных сигнальных групп увеличится.

Рассмотренный пример ММО с автономным и ортогональным по частоте излучением каждого передающего канала обладает еще одним преимуществом. Как известно, при использовании OFDM сигналов серьезной проблемой в случае большого количества несущих является значительная величина пик-фактора (отношение максимального уровня мощности к минимальному в процессе излучения OFDM-пакета). При организации ММО-канала с разбивкой всего пакета частот на ортогональные группы, количество которых равно количеству антенных элементов N , в каждом канале одновременно излучается в N раз меньший набор несущих по сравнению с традиционным способом. В результате существенно упрощаются требования к разрядности цифроаналоговых преобразователей, к линейности мгновенного динамического диапазона усилителей мощности и всего передающего тракта.

По сравнению с традиционным подходом, недостатком такого метода является понижение результирующей мощности сигналов по каждой несущей, поскольку в излучении сигнальных групп участвует только один из каналов, а не все одновременно.

Что касается требований к разрядности АЦП приемной ЦАР, то они в данном случае остаются неизменными вследствие суммирования всех сигналов в апертуре антенны.

Для придания ортогональных свойств сигналам разных элементов в передающей ЦАР вместо рассмотренного варианта OFDM-модуляции может использоваться кодирование сигналов во временной области, т. е. каждому из каналов

назначается свой уникальной код, ортогональный кодам остальных каналов. Конечно, при таком подходе можно было бы запрашивать каналы передающей ЦАР поочередно, перейдя фактически к системе типа MISO (Multiple-Input-Single-Output), но в результате существенно возрастает время передачи, и в целом нельзя гарантировать, что параметры каналов не изменятся за время сеанса. Кроме того, раздельная работа каналов на излучение не позволяет эффективно учесть их взаимное влияние, например путем внесения специальных групповых предискажений в сигналы, подлежащие передаче.

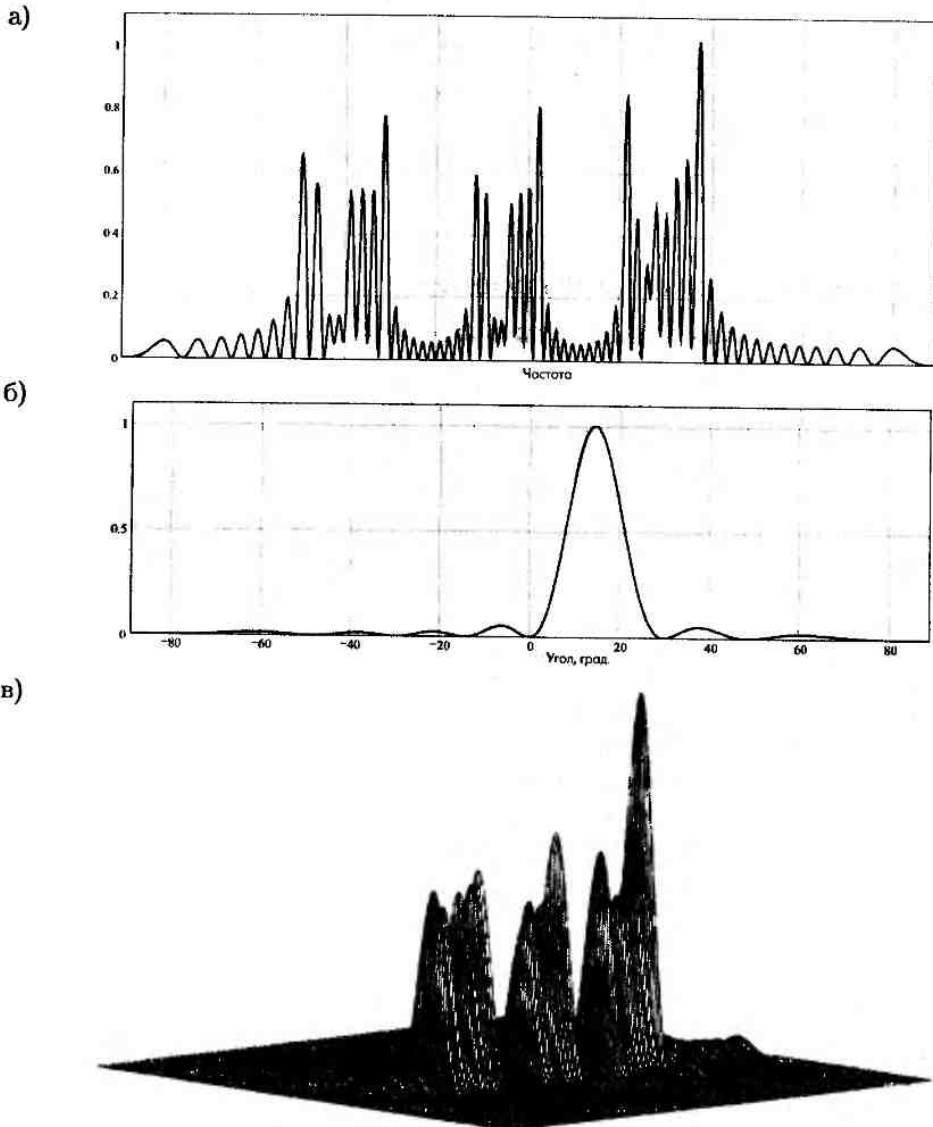


Рис. 9.73. Результаты моделирования работы ММО-системы

Для улучшения пространственного разрешения сигналов передающего сегмента ММО при отсутствии их ортогональности по каким-либо параметрам могут использоваться особенности естественного либо искусственно созданного ландшафта в интересах преднамеренной организации многолучевого распространения радиоволн и создания переотражений. В итоге на приемную решетку сигналы будут приходить с увеличенным угловым разнесением.

Теория обработки сигналов в ММО-системах переживает сегодня бурное развитие (этап становления). Появилось достаточно много методов обработки сигналов ММО, которые базируются на разложении корреляционной сигнальной матрицы на собственные вектора и собственные числа, при этом кодирование информации может быть заложено в величине собственных чисел. Существенно, что прогресс в развитии ММО-систем привел к появлению специализированных микросхем для такой обработки, резко удешевляющих производство соответствующего оборудования и создающих предпосылки для широкого внедрения ММО-технологии в средствах связи.

9.9 Особенности построения суперлинейных усилителей

Высокие требования, предъявляемые в ЦАР к величине мгновенного (без АРУ) линейного динамического диапазона (МЛЛД) аналоговых приемных каналов (например, при использовании многочастотных OFDM-сигналов), делают весьма проблематичной их реализацию на практике. Среди возможных путей решения этой проблемы важная роль отводится специализированным схемотехническим приемам, позволяющим при проектировании ультралинейных усилителей добиваться величины МЛЛД в 70 дБ и более. В числе указанных схемотехнических решений следует рассмотреть так называемые «двухтактные» или Push-Pull-усилители. Их типичная структурная схема приведена на рис. 9.74. Отличительной особенностью данной схемотехнической концепции является двухканальное противофазное усиление сигналов с последующим объединением откликов подканалов в выходном сумматоре. При этом происходит подавление высших паразитных гармоник в спектре сигнала. Для формирования противофазных сигналов на входе усилителя используется специальный трансформатор «расщепитель фазы».

Преимущества двухтактного усилителя следующие.

1. Четырехкратный прирост в величине входного импеданса устройства по сравнению с импедансом обычного несимметричного усилителя с той же самой выходной мощностью. В результате, двухтактная схема упрощает согласование.
2. Наличие «виртуальной» земли, что позволяет использовать большое количество компактных и более простых согласующих устройств.
3. Отсутствие в выходном сигнале четных гармоник $2F_1$, $2F_2$ и т. п., а также их комбинаций типа $F_2 - F_1$, $F_1 + F_2$ и т. д.

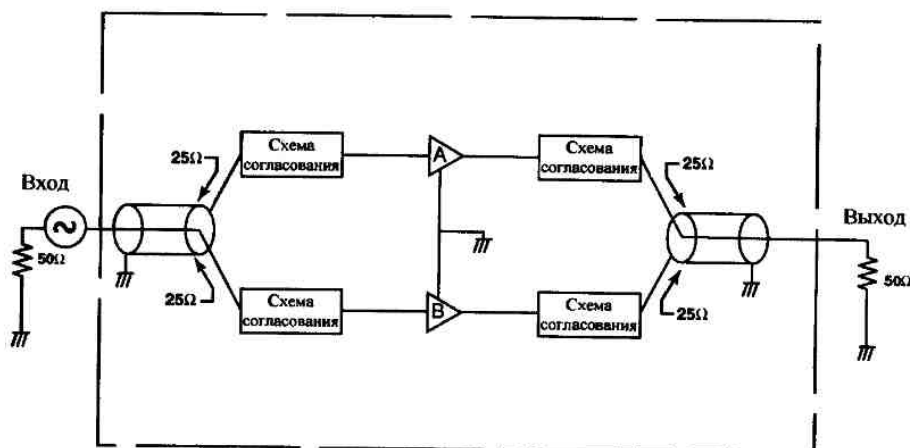


Рис. 9.74. Усилитель типа Push-Pull

Недостатки двухтактного усилителя следующие.

1. Использование симметричных трансформаторов в качестве противофазных расщепителей фазы не устраняет паразитных отражений мощности на входе и выходе устройства.
2. Изоляция между двумя противофазными частями усилителя теоретически составляет только 6 дБ. Такая недостаточная межэлементная развязка может вызывать проблемы неустойчивости.
3. Применение симметричных трансформаторов создает технологические сложности: вручную сделанный коаксиальный трансформатор прост в изготовлении только при использовании в экспериментальных лабораториях, но в серийном производстве это требует наличия дополнительной рабочей силы, что делает массовые поставки достаточно сложной задачей. Симметричные трансформаторы в виде микросборок хотя и доступны на рынке, но имеют более высокую стоимость и, как правило, занимают больше места, чем эквивалентные квадратурные расщепители фазы.

Традиционной конфигурацией усилителей для СВЧ-приложений является балансный усилитель (рис. 9.75). В нем используются 90-градусные разделитель сигналов (на входе) и объединитель (на выходе).

Преимущества балансного усилителя следующие.

1. Хорошая изоляция между двумя половинами устройства, что улучшает стабильность усилителя в широкой полосе частот.
2. Хорошая согласованность входа и выхода благодаря использованию 50-омной резисторной нагрузки, поглощающей возможные отражения сигналов.
3. Отсутствие гармоник типа $2F_1 + F_2$, $2F_2 + F_1$, $3F_1$, $3F_2$... и ослабление на 3 дБ комбинационных частот типа $F_1 - F_2$, $F_1 + F_2$, $2F_1$, $2F_2$...
4. Простота проектирования и интегрирования квадратурных расщепителей фазы.

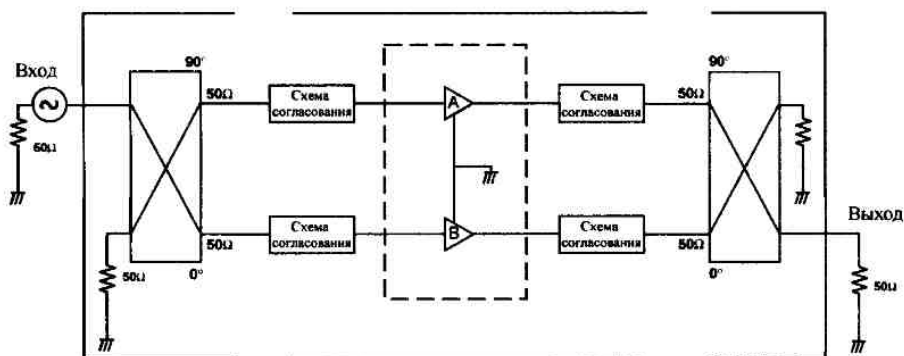


Рис. 9.75. Балансный усилитель

Недостатки конфигурации балансного (симметричного) усилителя следующие.

1. Требуется использование нагрузки 50 Ом. Резисторы большой мощности могут быть дороги и требуют надлежащего охлаждения.
2. Отсутствие виртуальной земли.

Сравнение характеристик двухтактного и балансного усилителей позволяет сделать следующие выводы.

1. В узкополосных (5–10%) приложениях показатели двух типов усилителей практически эквивалентны. Внешние согласующие цепи, разделители и объединители сигналов имеют схожие потери. Вследствие того, что четные гармоники находятся за пределами полосы пропускания согласующих цепей (внутренней и внешней) и симметричных трансформаторов, исключение этих гармоник в случае узкой полосы несущественно, и, следовательно, для узкой полосы двухтактная конфигурация не более эффективна, чем балансная. Линейность передаточной характеристики обоих типов усилителей в полосе частот, меньшей октавы по ширине, одинакова.
2. Двухтактная конфигурация предоставляет двукратное преимущество в коэффициенте трансформации импеданса. Это может упростить задачу согласования усилителей.
3. Балансные усилители имеют существенное преимущество во внешнем согласовании, они более устойчивы из-за хорошей изоляции между двумя каналами устройства.
4. Виртуальная земля в двухтактной конфигурации может рассматриваться как преимущество, поскольку она позволяет использовать подстроечные конденсаторы между двумя подканалами усилителя для его быстрой настройки.

Для иллюстрации незначительных различий между балансным и двухтактным усилителями можно сослаться на статистические данные табл. 9.6 [198],

которые соответствуют двум проектам усилителей, разработанным для диапазона 2,11–2,17 ГГц с использованием арсенидгаллиевых полевых транзисторов FLL1500IU-2С фирмы Fujitsu. Статистика была собрана путем усреднения характеристик пяти усилителей каждого типа. Параметры усилителей обеих схем достаточно подобны, за исключением внешнего согласования, которое лучше в случае балансной конфигурации.

Таблица 9.6. Сравнение балансного и двухтактного усилителей (статистические данные)

Параметры усилителей, измеренные при напряжении питания 12 В и потребляемом токе 4 А	Push-Pull-усилитель	Балансный усилитель
Коэффициент усиления (Gain)	12 дБ	11,8 дБ
Потери по входу (Input Return Loss)	13,4 дБ	20,2 дБ
Выходная мощность	51,9 дБм	52,0 дБм
Уровень третьей гармоники (IMD3) при выходной мощности 43 дБм	–36,4 дБ	–38,5 дБ
Коэффициент полезного действия	51%	54%

Среди основных путей достижения высокого динамического диапазона аналоговых приемников в последнее время все чаще рассматривают применение активных рекурсивных фильтров [151, 168] (рис. 9.76) в составе балансных схем суперлинейных усилителей (рис. 9.77).

В частности, именно данное техническое решение фигурирует в качестве предпочтительного в шведском проекте адаптивной многофункциональной ЦАП [151]. Соответствующий высокочастотный, малошумящий усилитель балансного типа обеспечивает мгновенный линейный динамический

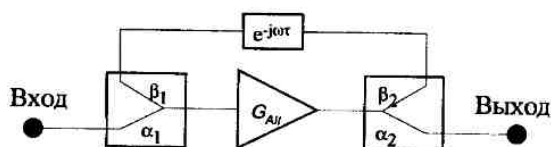


Рис. 9.76. Структура активного рекурсивного фильтра

диапазон более 64 дБ при полосе пропускания 18 МГц в диапазоне несущих 8–12 ГГц, и это не является пределом. Подтверждением тому могут служить серийно выпускаемые фирмой Mini-Circuit's балансные усилители серии ZRL (ZRL-700, ZRL-1200, ZRL-2300), структурная схема которых приведена на рис. 9.78. Согласно [172] их мгновенный динамический диапазон неискаженной интермодуляционными помехами передачи сигналов превышает 82 дБм (рис. 9.79) при коэффициенте усиления ниже 30 дБ. Указанные усилители могут рассматриваться в качестве реальных прототипов для создания ЦАП в интересах связных приложений.

Следует отметить, что, задавая требования к усилителям аналоговых приемников при их стыковке с АЦП, важно предусмотреть выполнение условия согласования СКО шумов с квантом АЦП в определенной пропорции. Как правило, их величины должны находиться в соотношении 1 : 1 (рис. 9.80), а еще лучше, чтобы на СКО шума приходилось несколько квантов АЦП. В последнем случае

когерентное накопление сигналов с «вытаскиванием» их из-под шума будет более эффективным.

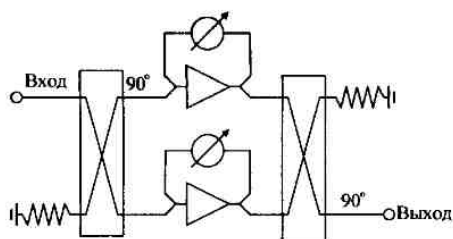


Рис. 9.77. Активные рекурсивные фильтры в составе балансного усилителя

Другое условие предполагает, чтобы компрессия усиления начиналась за пределами разрядной сетки АЦП (рис. 9.79), что достигается увеличением результирующего коэффициента усиления с помощью дополнительных выходных каскадов в аналоговом приемнике. В дальнейшем требуемое отношение сигнал-шум будет достигаться с помощью цифрового накопления сигналов путем их синфазного сложения по полотну решетки и когерентного накопления во времени, как это делается, например, в случае OFDM-сигналов.

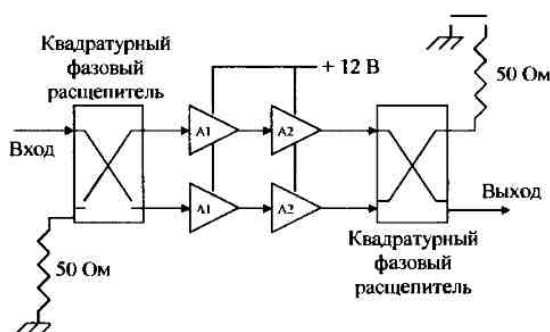


Рис. 9.78. Структурная схема балансных усилителей серии ZRL фирмы Mini-Circuit's

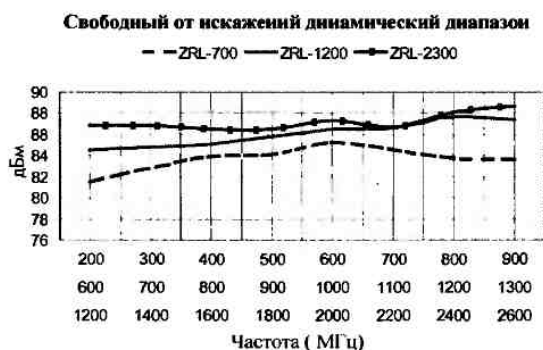


Рис. 9.79. Свободный от искажений мгновенный динамический диапазон усилителей серии ZRL фирмы Mini-Circuit's при полосе пропускания 1 МГц

В завершение следует отметить, что рассмотренные в данной главе схемотехнические аспекты построения систем широкополосной связи относятся к одной из ключевых и наиболее динамично развивающихся областей радиотехники. В дальнейшем заинтересованным читателям имеет смысл самостоятельно отслеживать

публикации по данной тематике, в том числе и в Интернете, чтобы однажды не быть застигнутыми врасплох очередным технологическим прорывом в технике связи.

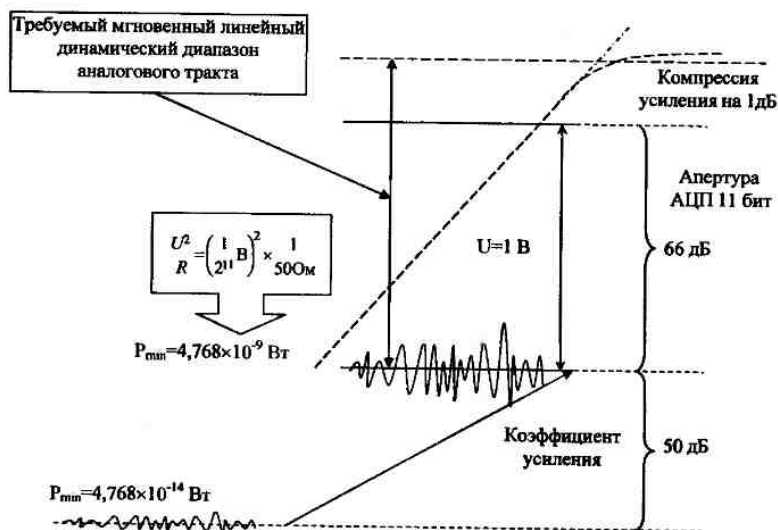


Рис. 9.80. Соотношение между мгновенным линейным динамическим диапазоном усилителя и его коэффициентом усиления

9.10 Построение беспроводных сетей передачи информации с использованием привязных аэростатов

Очевидным путем расширения области покрытия беспроводной сети является подъем базовой станции, с тем чтобы обеспечить прямую видимость для максимального числа пользователей. В последнее время в России, Европе, США и Японии прорабатываются концепции использования аэростатной техники в качестве несущих телекоммуникационных платформ или высотных платформ (HAPs — High Altitude Platforms). На это есть несколько причин. Аэростаты обеспечивают высотное положение передатчиков и ретрансляторов и способны заменить десятки и сотни типовых мачт. Стоимость аэростатной телекоммуникационной платформы на порядок ниже суммарных затрат на мачты-антенны. Качество передаваемого сигнала через высотную телекоммуникационную платформу аналогично по характеристикам спутниковому. С помощью аэростатных систем можно комплексно решать задачи телекоммуникационного и информационного обеспечения в условиях, где невозможно или нецелесообразно прокладывать кабель.

Есть еще одно важное преимущество аэростатных сетей, особенно актуальное в нашей стране. Количество высотных сооружений в г. Москве весьма ограничено, что характерно и для других городов России. Такие высотные сооружения превращаются в «сборище» разнообразных антенн, работающих в различных

диапазонах длин волн. Это приводит к серьезным проблемам электромагнитной совместимости и ухудшению связи в каждой из сетей.

В данном разделе проводится классификация существующих технологий создания региональных беспроводных сетей с помощью высотных платформ, описываются преимущества НАР-технологии на основе привязных аэростатов, проводится анализ структуры подобных систем и проблем, решаемых при их разработке. В качестве примера успешной реализации таких систем описывается российская аэростатная система БАРС, являющаяся элементом беспроводной региональной сети.

9.10.1 Классификация технологий высотных платформ

В настоящее время применяются следующие три технологии [207] для разработки высотных платформ для организации региональных беспроводных сетей.



Рис. 9.81. Высотная платформа для целей наблюдения

1. **Стратосферные дирижабли** длиной порядка 200 м (рис. 9.81), поднимаемые на высоту от 15 до 25 км. Эти рабочие высоты выбраны ввиду относительно низких скоростей ветра и турбулентности [207], а также отсутствия проблем, связанных с авианавигацией. Эти дирижабли должны работать автономно в течение долгого времени (до нескольких лет), используя энергию от солнечных батарей, и могут нести нагрузку в сотни килограмм при ожидаемой стабильности положения, необходимой для обеспечения обслуживания, примерно в пределах 1 км.

Среди проектов беспроводных телекоммуникационных сетей, использующих стратосферные дирижабли, следует назвать: сеть стратосферных дирижаблей SkyNet [148] (проект, возглавляемый Yokosuka Communications Research Laboratory, Япония), которая должна будет обеспечить покрытие большей части Японии; системы дирижаблей разработки Advanced Technologies Group (Bedford, UK) [115], SkyStation International (USA) [201], Lindstrand Balloons [171], The University of Stuttgart (Germany) [190] и др. Наряду с таким несомненным преимуществом, как весьма обширная зона покрытия, позволяющая реализовывать глобальные сети, данной технологии присущ целый ряд недостатков: длительные сроки разработки (несколько лет), высокая стоимость (десятки миллионов долларов), проблемы с удержанием станции и стабилизацией положения, эксплуатационные проблемы, а также необходимость в точной системе наведения в случае, если используется массив антенн для создания сотовых структур [207]. Высотных аэростатов, способных точно удерживать стабильное положение в пространстве, пока нет.

2. **Размещение базовых станций на пилотируемых или беспилотных самолетах.** Это легкие самолеты (рис. 9.82), значительно меньшие аэростатов, снабжаемые обычно энергией от солнечных батарей, должны лететь против ветра или по кругу ограниченного диаметра в целях позиционирования станции.

Самолеты беспилотные, имеют размах крыльев, как правило, не более 10 м, но могут нести полезную нагрузку до 100 кг. Их называют также солнечными аэродинамическими «платформами очень высокой долговечности» (HAVE — High Altitude Very Long Endurance) или «Гелиоплатформами» (HELIPLAT — Helios Platforms) [204].

Рис. 9.82. Экспериментальные солнечные самолеты, разработанные НАСА



Centurion (1998)



Pathfinder Plus (1998)

Основной проблемой этой технологии является необходимость накопления энергии для ночной работы. Автономность солнечных самолетов пока еще только изучается. Другие недостатки — примерно те же, что и в случае стратосферных платформ. Данная технология реализуется в таких проектах, как Helios plane (программы Pathfinder и Centurion американской компании AeroVironment) [116], HeliNet (финансируемый Европейским сообществом в рамках программы IST FP5) [204], HALO (проект Angel Technologies) [146].

3. Телекоммуникационные сети с использованием привязных аэростатов. Это наиболее простая и дешевая технология реализации высотных платформ, при которой в основном решается проблема удерживания станции. Кроме того, привязной трос позволяет легко осуществлять подачу питания на станции и доставлять данные к ней и от нее (путем оптоволокну, размещенного внутри троса). Основной проблемой этой технологии является авиатрафик, т.е. запуск подобных платформ возможен только в зонах, свободных от авианавигации, или в периоды, когда рабочая зона покрытия и рабочие высоты свободны от полетов. Одними из примеров успешной реализации таких систем являются:

- БАРС (русская аббревиатура Беспроводной Аэростатной РадиоСети), Москва, Россия [14, 184, 215]. Длина троса — до 1 км. Испытания проведены в 1999 году, система успешно функционирует, обеспечивая зону покрытия диаметром 50–70 км.
- ARC (Airborne Relay Communications) system, продукт Platform Wireless International [184]. Длина троса — до 4,6 км. Успешные испытания проведены в 2001 году в Сан-Диего, США. Цель проекта — развертывание аэростатной беспроводной сети над территорией Бразилии.
- SkyLINC [200]. В ходе проекта будет развернута сеть из 18 привязных аэростатов над территорией Великобритании. Каждый аэростат, удерживаемый тросом длиной до 1,5 км, обеспечит зону покрытия (соту) до 3,200 кв. км. Первая очередь данной сети будет запущена в начале 2004 году.

9.10.2 Преимущества беспроводных сетей на основе привязных аэростатов

Использование привязных аэростатов для создания телекоммуникационных сетей имеет следующие основные преимущества по сравнению с традиционными методами размещения базовых станций на высотных сооружениях. К ним относятся:

- резкое расширение зоны покрытия;
- низкая стоимость и краткие сроки реализации и развертывания;
- простота реконфигурации и масштабирования;
- идеальная приспособленность для поддержки широкополосных сервисов (например, цифрового телевидения, электронной рекламы и т. п.);
- уменьшение уровня помех от наземных радиоэлектронных средств (РЭС);
- экологичность метода построения радиосети передачи информации;
- многофункциональность (возможность использования для видеонаблюдения, мониторинга окружающей среды, телевидения, пейджинговой связи и т. д.);
- возможность мобильного и оперативного развертывания сети в условиях чрезвычайных ситуаций;
- обеспечение равных возможностей доступа в Интернет для всех пользователей в зоне покрытия.

Расширение зоны покрытия и уменьшение теневых зон. Использование привязных аэростатов с высотой подъема до 1000 м для размещения антенных устройств актуально не только в зонах с малой плотностью населения, но и в мегаполисах. Это объясняется тем, что используемое радиооборудование, обеспечивающее широкополосный доступ наземных абонентов к сетям передачи данных, работает в диапазоне частот от 2,5 ГГц и выше. Этот диапазон частот характерен тем, что информационные сигналы (радиосигналы) распространяются прямолинейно, практически без огибаний, поэтому наличие прямой видимости для такого оборудования приобретает первостепенное значение. Хотя современная OFDM-технология передачи данных делает возможным работу радиоустройств на отраженных сигналах, тем не менее такой канал связи получается нестабильным по производительности, и надежность его значительно хуже, нежели канала, образуемого в зоне прямой видимости. Размещение антенных устройств на указанной выше высоте позволяет минимизировать теневые зоны, обусловленные наличием высотных зданий в мегаполисах, и обеспечить подключение практически любых абонентов к мировой сети Internet (т. е. реализовать последнюю милю) по высокоскоростным каналам радиодоступа.

Кроме того, для еще большего расширения зоны покрытия и организации региональных и глобальных сетей может создаваться целая сеть аэростатов (пример — SkyLINC [200]). При этом базовые станции на высотных платформах могут

быть связаны между собой высокоскоростными оптическими линиями связи для организации опорной сети.

Уменьшение помех от наземных РЭС. При работе направленных антенн, которые используются в радиоаппаратуре базовой станции аэростата и наземных абонентов аэростатной радиосети, уменьшается мощность помех от наземных источников. Расчеты подтверждают уменьшение плотности мощности таких помех начиная с высоты 200 м. Уменьшение плотности мощности помех на высоте 800 м составляет 5–18 дБ.

Экологичность. При данной технологии источники излучения удаляются от мест работы и проживания людей, тем самым минимизируя возможные вредные воздействия электромагнитного излучения на природу и человека, по сравнению с наземными радиосетями. Кроме того, отсутствуют вредные выбросы при запуске аэростата с базовой станцией, по сравнению со спутниковыми системами.

Многофункциональность. Аэростатный носитель, используемый для создания телекоммуникационной сети, одновременно может быть использован для целого ряда назначений:

- видеонаблюдение за транспортными потоками, контроль государственных границ, обнаружение очагов пожаров и т. д.;
- мониторинг окружающей среды (сбор, обработка и передача информации с датчиков слежения за состоянием воздушной и водной сред);
- разворачивание новых дециметровых каналов телевидения.

9.10.3 Архитектура и проблемы реализации

Беспроводная коммуникационная сеть с использованием привязных аэростатов (рис. 9.83) включает следующие основные компоненты:

- гелиевый аэростатный носитель объемом не ниже 400 м³ и грузоподъемностью не ниже 50 кг полезной нагрузки;
- базовая станция с широкополосным радиооборудованием стандартов IEEE 802.11х (как в сети BAPC [14, 215]) или 802.16х [164] (например, оборудование с технологией W-OFDM типа Libra, как в сети SkyLINC [200]) и системой стабилизации положения антенн;
- привязной трос, обеспечивающий удержание аэростата, энергоснабжение и, в некоторых случаях, передачу данных по оптическому кабелю, расположенному внутри кевларового троса;
- лебедка для подъема и опускания аэростата;
- причальное устройство для удержания аэростата на земле.

При мобильном варианте реализации причальное устройство и лебедка располагаются в кузове специально оборудованного автомобиля.

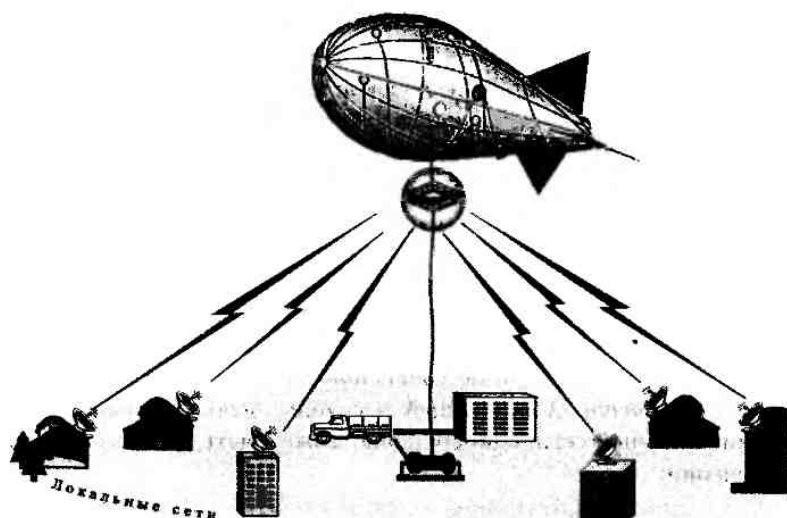


Рис. 9.83. Архитектура беспроводной сети на основе привязного аэростата

Однако при размещении базовой станции на привязном аэростате возникает целый ряд сложных технических проблем. Под действием ветровых нагрузок аэростат перемещается внутри конуса с вершиной в точке крепления привязного троса. Это приводит к необходимости создания системы стабилизации положения антенных устройств в пространстве (вертикальная и азимутальная стабилизация). В рамках разработки и создания в 1999 году в России (г. Москва) телекоммуникационной сети БАРС проблема стабилизации была решена. Вертикальная стабилизация положения антенн достигалась путем расположения каркаса крепления антенн на титановом сферическом подшипнике с демпфирующим устройством для гашения колебаний (рис. 9.84).

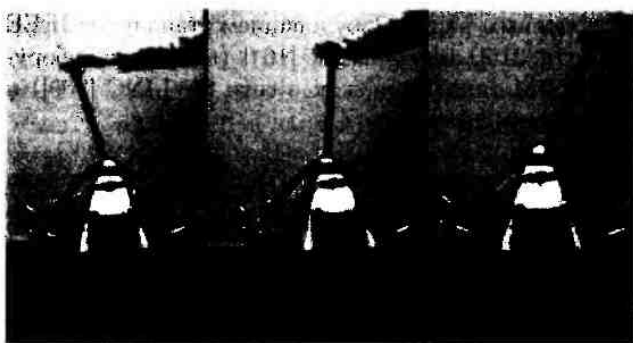


Рис. 9.84. Устройство стабилизации и базовая станция

Для азимутальной стабилизации разработан оригинальный и экономичный программно-технический комплекс, обеспечивающий точность стабилизации до

одного градуса, и с временем возврата платформы стабилизации к исходному направлению не более $0,25^\circ$. В состав комплекса входят: микропроцессор, электронный компас, блок коммутации, серверный двигатель, наземный терминал контроля управления и сбора статистики.

Бортовой и наземный программный комплекс включает в себя:

- комплекс программ слежения за стабилизирующими параметрами системы;
- комплекс программ управления исполнительными механизмами;
- комплекс программ оператора для наземной ЭВМ;
- комплекс программ сбора статистики;
- программный комплекс связи с наземным оператором.

Другой проблемой является проблема грозозащиты оборудования. Тривиальным решением этой проблемы, приводящим к временному прекращению работы радиосети, является опускание аэростата на землю при грозовом предупреждении. Альтернативным решением является встраивание в трос молниеотводящего элемента, что, однако, увеличивает вес троса. Имеются и другие подходы к решению проблемы грозозащиты.

Наконец, последней задачей для обеспечения нормального функционирования аэростатной базовой станции является обеспечение климатической стабилизации оборудования. Для решения этой проблемы оборудование базовой станции и платформа стабилизации закрываются радиопрозрачным или, если необходимо, оптически прозрачным обтекателем. Такое решение позволяет полностью убрать влияние ветровых нагрузок на платформу стабилизации, защитить оборудование от влияния атмосферной влаги, упростить температурную стабилизацию.

9.10.4 Проект БАРС

Перечисленные проблемы были успешно решены в рамках реализации проекта БАРС [14,210,215]. Беспроводная телекоммуникационная сеть БАРС была разработана и реализована Институтом проблем передачи информации РАН совместно с РНЦ «Курчатовский институт», Русским воздухоплавательным обществом «Авгуръ» и ракетно-космической корпорацией «Энергия». Указанная сеть обеспечивает подключение к Internet ряда образовательных и научных учреждений г. Москвы.

При работе над проектом было разработано специальное причальное устройство (рис. 9.85), включающее лебедку и особый кевларовый кабель-трос, комбинированный с оптоволоконным и электрическим кабелями. С помощью этих приспособлений можно плавно поднимать и опускать аэростат. Кроме того, по кабель-тросу осуществляется подача питания на аэростат и связь базовой станции с землей по оптоволоконному каналу.

Кабель-трос состоит из следующих основных элементов (рис. 9.86):

- направляющего стержня (6), вокруг которого навиваются волокна СВМ (4),
- синтетической оболочки (2), закрывающей СВМ волокна,

- трех медных многожильных изолированных проводников (3), общим сечением $0,5 \text{ мм}^2$,
- одного оптико-волоконного кабеля (5) толщиной $0,3 \text{ мм}$, в котором находятся два светлых кварцевых оптических волокна в изоляции по 250 мкм каждый,
- внешней климатической защиты (1).

Рис. 9.85. Причалное устройство

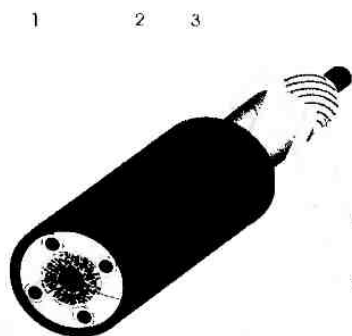


Рис. 9.86. Структура кабель-троса

Специфической проблемой оказалась задача стабилизация платформы, на которой крепилось передающее оборудование. Была реализована конструкция, показанная на рис. 9.84.

На шаровом шарнире свободно вращается «юбочка», к которой крепится платформа с оборудованием. При возможных отклонениях от вертикали платформа остается в горизонтальной плоскости. Об азимутальной стабилизации в этой реализации проекта беспокоиться не приходилось, так как была использована всенаправленная передающая антенна.

Кроме того, были созданы конструкции для размещения базовой станции, специальных антенн, устройства стабилизации, а также конструкции обтекателя и платформы для установ-

ки оборудования. Обтекатель (рис. 9.87) представляет собой радиопрозрачный кожух, защищающий оборудование базовой станции и системы стабилизации от климатических воздействий.

В качестве носителя платформы был взят малообъемный аэростат Au-17 (рис. 9.88). Телекоммуникационное оборудование комплекса представлено радиоMODEMами Cisco WGB340/WGB350 и Cisco BR340/BR350, а также всенаправленной антенной, антенными усилителями и маршрутизаторами.

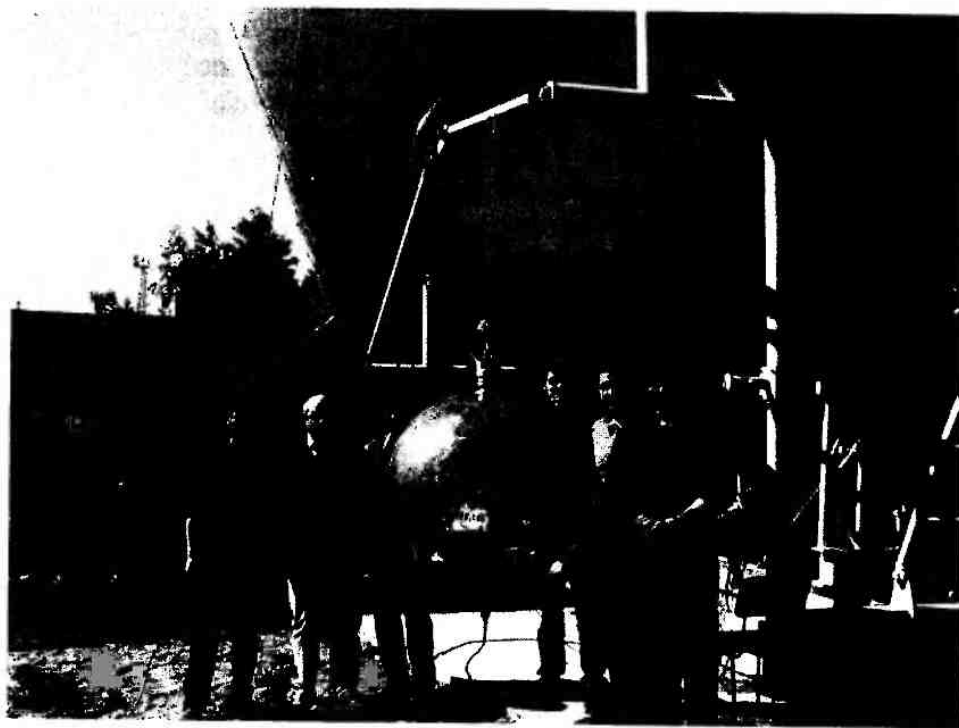


Рис. 9.87. Базовая станция и система стабилизации, закрытые радиопрозрачным кожом

Абонентские станции работали на базе радиомодема ARLAN.

Максимальная грузоподъемность аэростатной установки – 120 кг, что позволяет реализовывать телекоммуникационные проекты (например, базовая станция стандарта CDMA весит 54 кг).

Дополнительно, если позволяет масса, на платформе системы БАРС может быть установлено оборудование специального назначения: военные системы обнаружения; камеры видеонаблюдения, с помощью которых осуществляется мониторинг местности. В том числе могут быть установлены и параллельно функционирующие станции, работающие на разных частотах.

Передача сигнала через БАРС осуществляется в диапазоне 2.4 ГГц в рамках стандарта IEEE 802.11b.

Технические данные привязного аэростата Аи-17 «БАРС»

Объем оболочки для заполнения гелием, м ³	450
Длина оболочки, м	18,9
Максимальная скорость ветра, м/с	
при подъеме	20
на рабочей высоте	25

Диаметр оболочки, м	8,6
Масса, кг	155
Рабочая высота с полезной нагрузкой, м	1000
Максимальная полезная нагрузка, кг	120
Суммарная мощность всех потребителей, включая полезную нагрузку, кВт, не более	2,5
Длина кабель-троса, м	1200
Диаметр кабель-троса, мм	8



Рис. 9.88. Аэростат

Литература

- [1] Альтшуллер Г. С. *Найти идею: Введение в теорию решения изобретательских задач*. Новосибирск: Наука. Сиб. отд., 1986.
- [2] Артамонов А.А., Протопопов Л.Н., Варгузин В.А. Первый опыт цифрового наземного ТЕЛЕвещания в России// *Телемультимедиа*. 2000. № 3.
- [3] Баканов А.С., Вишневский В.М., Ляхов А.И. Метод оценки показателей производительности беспроводных сетей с централизованным управлением// *Автоматика и телемеханика*. 2000. № 4. С. 97–105.
- [4] Банкет В. Л., Дорофеев В. М. *Цифровые методы в спутниковой связи*. М.: Радио и связь, 1988.
- [5] Белов А. Полезные мелочи современной телефонии// *Электроника: НТБ*. 1999. № 2.
- [6] Берлекэмп Э. *Алгебраическая теория кодирования*. М.: Мир, 1971.
- [7] Блох Э.Л., Зяблов В.В. *Обобщенные каскадные коды*. М.: Связь, 1976.
- [8] Богуславский Л.Б., Ляхов А.И. Оценка производительности распределенных информационно-вычислительных систем архитектуры «КЛИЕНТ–СЕРВЕР»// *Автоматика и телемеханика*. 1995. № 9. С. 160–175.
- [9] Варганов М.Е., Зиновьев Ю.С., Астанин Л.Ю. и др. *Радиолокационные характеристики летательных аппаратов*. М.: Радио и связь, 1985.
- [10] Варгузин В., Артамонов А. Сравнительные характеристики европейского и американского стандартов цифрового наземного телевидения// *Теле-Спутник*. 1999. № 11.
- [11] Веселов Г.И., Егоров Е.Н., Алехин Ю.Н. и др. *Микроэлектронные устройства СВЧ: Учеб. пособие*. М.: Высшая школа, 1988.
- [12] Витерби А. Д., Омура Д. К. *Принципы цифровой связи и кодирования*. М.: Радио и связь, 1982.
- [13] Вишневский В.М. Беспроводные сети широкополосного доступа к ресурсам Интернета// *Электросвязь*. 2000. № 10. С. 9–13.
- [14] Вишневский В.М. *Теоретические основы проектирования компьютерных сетей*. М.: Техносфера, 2003.
- [15] Вишневский В.М. Методы, алгоритмы и программно-технические средства реализации широкополосных беспроводных сетей// *Second International conference «Information and Telecommunication Technologies in Intelligent Systems», Barcelona (Spain). Proceedings*. М.: МГИЭМ, 2004. С. 6–13.

- [16] Вишневский В.М., Воробьев В.М., Семушкина С.Г. Об одном подходе к решению проблемы «информационного неравенства» российских регионов// *Труды Международного семинара «Распределенные компьютерные и телекоммуникационные сети. Теория и приложения (DCCN-2003)»*. М.: Техносфера. 2003, Т. 2. С. 142–146.
- [17] Вишневский В. М., Гузаков Н. Н. Разработка мультилучевых антенных систем для базовых станций цифровой сети множественного доступа// *Second International conference «Information and Telecommunication Technologies in Intelligent Systems», Barcelona (Spain). Proceedings*. М.: МГИЭМ, 2004. С. 13–18.
- [18] Вишневский В.М., Гузаков Н.Н., Лаконцев Д.В. Беспроводная радиоэлектронная система «Рапира»// *Электроника: НТБ*. 2005. № 1.
- [19] Вишневский В.М., Гузаков Н.Н., Маднев Д.Н. Разработка серии отечественных радиоустройств для комплектации широкополосных беспроводных региональных сетей передачи данных, голоса и видеоинформации// *Труды Международного семинара «Распределенные компьютерные и телекоммуникационные сети. Теория и приложения (DCCN-2003)»*. М.: Техносфера, 2003, Т. 2. С. 202–210.
- [20] Вишневский В.М., Дмитриев В.П., Жданов В.С. *Основы передачи информации в вычислительных системах и сетях: Учебное пособие*. М.: МГИЭМ, 1998.
- [21] Вишневский В.М., Ляхов А.И. Оценка производительности беспроводной сети в условиях помех// *Автоматика и телемеханика*. 2000. № 12. С. 87–103.
- [22] Вишневский В.М., Ляхов А.И. Оценка пропускной способности локальной беспроводной сети при высокой нагрузке и помехах. *Автоматика и телемеханика*, 2001, № 8. С. 81–96.
- [23] Вишневский В.М., Ляхов А.И., Гузаков Н.Н. Оценка максимальной производительности беспроводного доступа в Интернет// *Автоматика и телемеханика*. 2004. № 9. С. 52–70.
- [24] Вишневский В.М., Ляхов А.И., Терещенко Б.Н. Моделирование беспроводных сетей с децентрализованным управлением// *Автоматика и телемеханика*. 1999. № 6. С. 88–99.
- [25] Вишневский В.М., Ляхов А.И., Терещенко Б.Н. и др. Региональные беспроводные сети передачи данных на базе протокола RADIO-ETHERNET: состояние, моделирование, примеры реализации// *Информационные процессы*. 2001, Т. 1, № 1. С. 10–32.
- [26] Возенкрафт Д., Рейффен Б. *Последовательное декодирование*. М.: Иностранная литература, 1963.
- [27] Галлагер Р. Коды с малой плотностью проверок на четность// *Теория кодирования*. М.: Мир, 1966. С. 139–165.
- [28] Галлагер Р. *Теория информации и надежная связь*/ Пер. с англ. под ред. М. С. Пинскера и Б. С. Цыбакова. М.: Советское радио, 1974.
- [29] Гинзбург В.В. Многомерные сигналы для непрерывного канала// *Проблемы передачи информации*. 1984. Т. 21. № 1. С. 14–27.
- [30] Гольдштейн Б.С. *Сигнализация в сетях связи. Т. 1*. М.: Радио и связь, 1998.

- [31] Гольдштейн Б.С. *Протоколы сети доступа. Т. 2.* М.: Радио и связь, 1999.
- [32] Гольшко А. Большая семья WLL// *Мир связи — Connect!* 1998. № 5.
- [33] Гольшко А. В. Телевидение: от кабельного к эфирному и далее...// *Сети и системы связи.* 1998. № 3.
- [34] Гольшко А., Морозюк В., Муратов Е. Вестница связи// *Вестник связи.* 2000. № 6.
- [35] Горобець Ю.І., Кучко А.М. *Вступ до фізики фрактальних структур: Навч. посібн.* Київ, 2000.
- [36] Драбкин А.Л., Зузенко В.Л., Кислов А.Г. *Антенно-фидерные устройства.* М.: Советское радио, 1974.
- [37] Зигангиров К. Ш. *Процедуры последовательного декодирования.* М.: Связь, 1974.
- [38] Зиновьев В.А., Зяблов В.В., Портной С.Л. *Каскадные методы построения и декодирования кодов в Евклидовом пространстве.* М.: Препринт ИППИ АН СССР, 1987.
- [39] Зяблов В.В., Коробков Д.Л., Портной С.Л. *Высокоскоростная передача сообщений по реальным каналам.* М.: Радио и связь, 1990.
- [40] Зяблов В.В., Коробков Д.Л., Портной С.Л. *Каскадные методы кодирования и синтеза сигналов в каналах с межсимвольной интерференцией.* М.: Препринт ИППИ АН СССР, 1988.
- [41] Касами Т., Токура Н., Ивадари Е., Инагаки Я. *Теория кодирования.* М.: Мир, 1978.
- [42] Кириллов В., Михеев П. Расстояния на миг сократив... (обзор геостационарных спутниковых систем связи, часть 1)// *ТЕЛЕ-Спутник.* 1997. № 5(19).
- [43] Кларк Д., Кейн Д. *Кодирование с исправлением ошибок в системах цифровой связи.* М.: Радио и связь, 1987.
- [44] Конкурирующие стандарты цифрового телевизионного вещания// *Электроника: НТБ.* 2001. № 1. С. 17–19.
- [45] Константинов Е. Реалии широкополосного доступа, или о том, чем Россия пока не избалована// *Вестник связи.* № 6. 2004.
- [46] *Концепция развития связи РФ/ Под ред. В.Б. Булгака.* 1996.
- [47] Корбанский И.Н. *Антенны.: Учебное пособие для вузов.* М.: Энергия, 1973.
- [48] Коростелев А.А. *Пространственно-временная теория радиосистем: Учеб. пособие для вузов.* М.: Радио и связь, 1987.
- [49] Котельников В.А. О пропускной способности эфира и проволоки в электросвязи// *Материалы к I Всесоюзному съезду по вопросам технической реконструкции дела связи и развития слаботочной промышленности.* М.: Всесоюз. энерг. комитеТ. Изд. Упр-ия связи РКК, 1933.
- [50] Кузнецов Н.А., Вишневский В.М., Гоев А.И., Дмитриев В.П. Беспроводные оптоэлектронные системы передачи информации// *ВКСС connect.* 2001. № 5. С. 19–23.

- [51] Кукк К.И. Цифровое телевидение в России// *Электроника: НТБ*. 2001. № 3.
- [52] Кухарев В. Перспективы развития и преимущества DVB-T в России и странах ближнего зарубежья// *www.nat.ru*
- [53] Лидл Р., Нидеррайтер Г. *Конечные поля*. Т. 1, 2. М.: Мир, 1988.
- [54] Липатов А., Варукина Л. Обеспечение информационной безопасности в системах DECT// *Технологии и средства связи*. 1999. № 5.
- [55] Ляхов А.И. Асимптотический анализ моделей иерархических локальных сетей с многопроцессорными серверами// *Автоматика и телемеханика*. 1998. № 12. С. 82–93.
- [56] Ляхов А.И., Мацнев Д.Н. Исследование функционирования городской беспроводной сети передачи данных// *Электросвязь*. 2003. № 6. С. 40–43.
- [57] Ляхов А.И., Мацнев Д.Н., Лаконцев Д.В., Шелихов О.Н. Сбор и анализ характеристик функционирования действующей беспроводной сети на основе протокола IEEE 802.11// *VIII международная конференция по информационным сетям, системам и технологиям (МКИССиТ-2002)*. СПб. 2002. С. 225–235.
- [58] Майника Э. *Алгоритмы оптимизации на сетях и графах*. Пер. с англ. М.: Мир, 1984.
- [59] Мак-Вильямс Ф.Дж., Слоэн Н.Дж.А. *Теория кодов, исправляющих ошибки*. М.: Связь, 1979.
- [60] Марпл-мл. С. Л. *Цифровой спектральный анализ и его приложения*. Пер. с англ. М.: Мир. 1990.
- [61] Материалы компании АНА// *www.aha.com*.
- [62] Материалы компании Алварион// *www.alvarion.com, www.alvarion.ru*.
- [63] Материалы компании Седиком// *www.cedicom.ru*.
- [64] Материалы сайта *www.turbocodes.ru*.
- [65] Мейтин М. Bluetooth: устройства всех стран, соединяйтесь! Без проводов// *Электроника: НТБ*. 2000. № 5.
- [66] Мейтин М. MPEG как предвестник эры интерактивного телевидения. *Электроника: НТБ*, 2001, № 4.
- [67] Месси Дж. *Пороговое декодирование*. М.: Мир, 1966.
- [68] Невдяев Л. CDMA: технологии доступа// *Сети*. 2000. № 6.
- [69] Нивников Д. BreezeNET-PRO// *PC Magazine Russian Edition*. 1998. № 5. С. 36–38.
- [70] Новаковский С. Последние данные о внедрении телевидения высокой четкости (стандарт ATCS) в США. 625. 2000. № 1.
- [71] *Pat. 3662316 США*. Short Base-Band Pulse Receiver/Kenneth W.Robbins. Приоритет 12.03.71.

- [72] Пат. 3728632 США. Transmission and Reception System for Generation and Receiving Base-Band Duration Pulse Signals without Distortion for Short Base-Band Pulse Communication System/Gerald F.Ross. Приоритет 12.03.71.
- [73] Пат. 3806795 США. Geophysical Subveying System Employing Electromagnetic Imulses/Rexford M.Morey. Приоритет 3.01.72.
- [74] Пат. 4641317 США. Spread Spectrum Radio Transmission System/Larry W. Fullerton. Приоритет 3.12.84.
- [75] Пат. 5677927 США. Ultrawide-Band Communication System and Method/Larry W. Fullerton, Ivan A. Cowie. Приоритет 20.09.94.
- [76] Пат. 5687169 США. Full Duplex Ultrawide-Band Communication System and Method/Larry W. Fullerton. Приоритет 27.04.95.
- [77] Петерсон У., Уэлдон Э. Коды, исправляющие ошибки. М.: Мир, 1976.
- [78] Портной С.Л. Характеристики систем модуляции и кодирования с точки зрения каскадных кодов// Проблемы передачи информации. 1985. Т. 20, № 1. С. 28–46.
- [79] Сазонов Д.М. Антенны и устройства СВЧ. М.: Высшая школа, 1988.
- [80] Семенов Ю.А. Протоколы и ресурсы Internet. М.: Радио и связь, 1996.
- [81] Сканирующие антенные системы СВЧ. Пер. с англ.// Под ред. Г.Т. Маркова, А.Ф. Чаплина. М.: Советское радио, 1966.
- [82] Скородумов А.И. Современное состояние и перспективы развития подвижной связи. Доклад на семинаре «Современное состояние и перспективы развития подвижной связи», 4 декабря 2003 года// www.sotovik.ru
- [83] Слюсар В.И. Ультразвуковая техника на пороге третьего тысячелетия// Электроника: НТБ. 1999. № 5. С. 50–53.
- [84] Слюсар В.И. Идеология построения мультистандартных базовых станций перспективных систем связи// Радиозлектроника (Изв. высш. учеб. заведений). 2001. № 4. С. 3–12.
- [85] Слюсар В.И. Синтез алгоритмов измерения дальности М источников при дополнительном стробировании отсчетов АЦП// Радиозлектроника (Изв. высш. учеб. заведений). 1996. № 5. С. 55–62.
- [86] Слюсар В.И. Схемотехника цифрового диаграммообразования. Модульные решения// Электроника: НТБ. 2002. № 1. С. 46–52.
- [87] Слюсар В.И. Цифровое формирование луча в системах связи: будущее рождается сегодня// Электроника: НТБ. 2001. № 1. С. 6–12.
- [88] Слюсар В.И. Цифровые антенные решетки: будущее радиолокации// Электроника: НТБ. 2001. № 3. С. 42–46.
- [89] Слюсар В.И., Титов И.В. Метод коррекции характеристик передающих каналов активной ЦАР// Известия вузов. Сер. Радиозлектроника. 2004. Т. 47. № 8. С. 14–20.
- [90] Слюсар В.И. Коррекция характеристик приемных каналов цифровой антенной решетки по контрольному источнику в ближней зоне// Радиозлектроника (Изв. высш. учеб. заведений). 2003. № 1. С. 44–52.

- [91] Слюсарь В.И. Фаббер-технологии: сам себе конструктор и фабрикант// *Конструктор*. 2002. № 1. С. 5–7.
- [92] *Справочник по антенной технике: В 5 т. Т. 1*// Л.Д. Бахрах, Л.С. Бенинсон и др.; Под ред. Я.Н. Фельда, Е.Г. Зелкина. М.: ИПРЖР, 1997.
- [93] Столлингс В. *Беспроводные линии связи и сети. Пер. с англ.* М.: Издательский дом «Вильямс», 2003.
- [94] Федоров П.Н. Мачты-деревья// *Конструктор*. 2001. № 2. С. 18.
- [95] Финк Л.М. *Теория передачи дискретных сообщений*. М.: Совесткое радио, 1963.
- [96] Форни Г. Д. *Каскадные коды*. М.: Мир, 1970.
- [97] Фрадин А.З., Рыжков Е.В. *Измерения параметров антенно-фидерных устройств*. М.: Связь, 1972.
- [98] Чернов Б. Что DECT грядущий нам готовит?// *Электроника: НТБ*. 2000. № 1. С. 8–14.
- [99] Черток Б.Е. *Ракеты и люди. Горячие дни холодной войны*. М.: Машиностроение, 2002.
- [100] Шахнович И. *Современные технологии беспроводной связи*. М.: Техносфера, 2004.
- [101] Шахнович И. Грядет тотальная МОБИЛизация// *Электроника: НТБ*. 2001. № 2. С. 32–38.
- [102] Шахнович И. Сотовый телефон третьего поколения// *Электроника: НТБ*. 2001. № 3. С. 20–23.
- [103] Шахнович И. Сверхширокополосная связь. Второе рождение// *Электроника: НТБ*. 2001. № 4. С. 8–15.
- [104] Шахнович И. Цифровое радиовещание — последний штрих цифрового портрета?// *Электроника: НТБ*. 2002. № 3. С. 8–15.
- [105] Шахнович И. Беспроводные локальные сети. Анатомия стандартов IEEE 802.11// *Электроника: НТБ*. 2003. № 1. С. 38–48.
- [106] Шахнович И. Беспроводные локальные сети IEEE 802.11 — G дали!// *Электроника: НТБ*. 2003. № 4. С. 36–39.
- [107] Шахнович И. Сети городского масштаба: решения рабочей группы IEEE 802.16 — в жизнь!// *Электроника: НТБ*. 2003. № 8. С. 50–56.
- [108] Шахнович И. Персональные беспроводные сети стандартов IEEE 802.15.3 и 802.15.4// *Электроника: НТБ*. 2004. № 6. С. 32–36.
- [109] Шахнович И. Стандарт широкополосного доступа IEEE 802.16 для диапазонов ниже 11 ГГц// *Электроника: НТБ*. 2005. № 1. С. 8–14.
- [110] Шахнович И. Стандарт широкополосного доступа IEEE 802.16. Режим OFDMA и поддержка адаптивных антенных систем// *Электроника: НТБ*. 2005. № 2. С. 46–52.
- [111] Шеннон К.Э. *Работы по теории информации и кибернетике*. М.: Иностранная литература, 1963.

- [112] Щербак Н. Сверхширокополосная радиолокация. Что это такое// *Электроника: НТБ*. 2002. № 3. С. 38–46.
- [113] Элайес П. Кодирование для двух каналов с шумами// *Теория передачи сообщений*. М.: Иностранная литература, 1957. С. 114–138.
- [114] Яноши Л. *Теория и практика обработки результатов измерений*. 2-е изд. Пер. с англ. М.: Мир, 1968.
- [115] Advanced Technologies Group// <http://www.airship.com>
- [116] AeroVironment// <http://www.aerovironment.com>
- [117] AirForce-BR100-R-11.14.02// *Broadcom*, 2002.
- [118] Alamouti, S.M. A Simple Transmit Diversity Technique for Wireless Communications// *IEEE Journal on Select Areas in Communications*. Oct. 1998. Vol. 16, No. 8. P. 1451–1458.
- [119] Anastasi G. and Lenzini L. QoS Provided by the IEEE 802.11 Wireless LAN to Advanced Data Applications: a Simulation Analysis// *Wireless Networks*. 2000. Vol. 6, No. 2. P. 99–108.
- [120] Anderson D.S., Drocella Ed.F., Jones S.K., Mark A.S. Assessment of Compatibility between Ultrawideband (UWB) Systems and Global Position System (GPS) Receivers// *NTIA Special Publication-01-45*. U.S. Department of Commerce. February 2001.
- [121] Arensman R. Intel, 11g fuel WLAN boom// *Electronic Business*, 5/1/2003.
- [122] ArrayComm Homepage// <http://www.arraycom.com/>
- [123] Barrett T.W. History of UltraWideBand (UWB) Radar & Communications: Pioneers and Innovators// *Progress in Electromagnetics Symposium 2000. July 2000*.
- [124] Barth P. A Davis-Putnam Based Enumeration Algorithm for Linear Pseudo-Boolean Optimization// <http://www.mpi-sb.mpg.de/units/ag2/software/opbdf/>
- [125] Basagni S., Conti M., Giordano S., Stojmenovich I. *Mobile Ad Hoc Networking*. John Wiley & Sons, 2004.
- [126] Bennett C.L., Ross G.F. Time-Domain Electromagnetics and Its Applications// *Proceedings of the IEEE*. 1978. Vol. 66. No. 3.
- [127] Berrou C., Glavieux A., Thitimajshima P. Near Shannon limit error-correcting coding and decoding: Turbo-codes// *ICC'93, Geneva, Switzerland*. May 93. P. 1064–1070.
- [128] Best S.R. The Koch Fractal Monopole Antenna: The Significance of Fractal Geometry in Determining Antenna Performance// *Proceedings of the 2001 Antenna Applications Symposium*. Allerton Park Monticello, Illinois, September 19–21, 2000// <http://www.ecs.umass.edu/ece/allerton/papers2001/2001-p194.pdf>
- [129] Bhoopalam Bh. MIMO Channel Capacity and space time coding// http://www.ece.drexel.edu/courses/ECE-T612/Talks/B_Bhoopalam.pdf
- [130] Bianchi G. Performance Analysis of the IEEE 802.11 Distributed Coordination Function// *IEEE Journal on Selected Areas in Communications* 18(3) (March 2000). P. 535–547.

- [131] Bing B. Broadband Wireless Access (*Kluwer International Series in Engineering and Computer Science*, 578). Kluwer Academic Publishers, 2000.
- [132] Bing B. High-Speed Wireless Atm and Lans (*Artech House Mobile Communications Library*). Artech House Publishers, 2000.
- [133] Bing B. Wireless Local Area Networks: The New Wireless Revolution. *Wiley-Interscience*, 2002.
- [134] Bononi L., Conti M., Donatiello L. Design and Performance Evaluation of Distributed Contention Control (DCC) Mechanism for IEEE 802.11 Wireless Local Area Network// *J. Parallel Distrib. Comput.*, 2000. Vol. 60, No. 4.
- [135] Bruno R., Conti M., Gregori E. Bluetooth: Architecture, Protocols and Scheduling Algorithms// *Cluster Computing*. 2002. Vol. 5, P. 117–131.
- [136] Bruno R., Conti M., Gregori E. A Simple Protocol for the Dynamic Tuning of the Backoff Mechanism// *IEEE 802.11 Networks. Computer Networks*. 2001. Vol. 37. No. 1. P. 33–44.
- [137] Calí F., Conti M., Gregory E. Dynamic Tuning of the IEEE 802.11 Protocol to Achieve a Theoretical Throughput Limit// *IEEE/ACM Transactions on Networking*, December 2000. No. 8(6). P. 785–799.
- [138] Carrier Sense Multiple Access with Collision Detection (CSMA CD) access method and physical layer specifications. ANSI IEEE Standard 802.3// *IEEE Press*. 1996.
- [139] Causley A.J. Design of Conformal Antennas for Telephone Handsets// *Bachelor of Engineering Honours Thesis*. The University of Queensland, 2002// <http://innovexpo.itee.uq.edu.au/2002/projects/s354019/thesis.pdf>
- [140] Chen Z.N., Chia M.Y.W. A Circular planar inverted-L antenna with vertical ground plane// *Microwave and Optical Technology Letters*. November 2002. Vol. 35. No. 4. P. 315–317// <http://www1.i2r.a-star.edu.sg/~chenzn/MOTL-1102.pdf>
- [141] Cheng L., Marsic I. Piecewise Network Awareness Service for Wireless/Mobile Pervasive Computing// *Mobile Networks and Applications (MONET)*. 2002. Vol. 7. No. 4. P. 269–278.
- [142] Chhaya H.S., Gupta S. Performance Modeling of Asynchronous Data Transfer Methods of IEEE 802.11 MAC Protocol// *Wireless Networks* 3(3). March 1997. P. 217–234.
- [143] Choi S. PCF vs. DCF: Limitations and Trends// *IEEE 802.11-01* 154. Jan. 2001.
- [144] Ciais P., Luxey C., Diallo A., Staraj R., Kossias G. Design of Internal Multiband Antennas for Mobile Phone and WLAN Standards// *Joint COST 273 284 Workshop on Antennas and Related System Aspects in Wireless Communications*, Chalmers University of Technology Gothenburg, Sweden, 2004// http://www.s2.chalmers.se/costworkshop/workshop_papers/126.pdf
- [145] Cohen N. U.S. Patent № 6140975. H01Q 1/48. Fractal Antenna Ground Counterpoise, Ground Planes and Loading Elements. Oct. 31, 2000.

- [146] Colella N.J., Martin J.N., Akyildiz I.F. The HALO network// *IEEE Commun. Mag.* June 2000. P. 142–148.
- [147] Collins B.C., Iellici D., Kingsley S.P., O'Keefe S.G., Ollitervo P.J., Zhang S. Antenna Ltd. A hybrid ceramic quadband antenna for handset applications// <http://www.antenna.com/media/papers/IEEE%20Symposium%20Shanghai.pdf>
- [148] Communications Research Laboratory, Japan// <http://www2.crl.go.jp/mt/b181/research/spf/index-e.html>
- [149] Conti M., Gregori E., Lenzini. L. *Metropolitan Area Networks*. Springer-Verlag Telos, 1997.
- [150] Cummings N.P. *Low Profile Integrated GPS and Cellular Antenna*. Master Thesis, Blacksburg: Virginia Polytechnic Institute, 2001// <http://scholar.lib.vt.edu/theses/available/etd-11132001-145613/unrestricted/etd.pdf>
- [151] Danestig M., Malmqvist R., Ouacha A. Design Considerations for Active MMIC Filters in Future Adaptive Radar Antennas// *EuMC 2000. Workshop WS8*.
- [152] DECT baseband controller PCD5091// *Data sheet, Philips Semiconductor*.
- [153] Dornan A., Dornan A. *The Essential Guide to Wireless Communications Applications 2nd Ed*. Prentice Hall PTR, 2002.
- [154] Draft New Recommendation ITU-R BS. System for Digital Sound Broadcasting in the Broadcasting Bands Below 30 Mhz// *ITU, Document 6/63-E*, 25 October 2000.
- [155] ETS 300 401. Radio Broadcasting Systems; Digital Audio Broadcasting (DAB) to mobile, portable and fixed receivers// *European Telecommunications Standards Institute*, 2001.
- [156] Fox B. Has the United States got it wrong with digital terrestrial TV? Digital TV Rollout// *IEEE Spectrum Online*. 2001. February.
- [157] GPSS World// www.minutemansoftware.com
- [158] Gianvittorio J. *Fractal Antennas: Design, Characterization, and Applications*. University of California, Los Angeles, 2000// http://www.antlab.ee.ucla.edu/~johng/fractals/master_thesis.pdf
- [159] Gilbert E.N. Capacity of a Burst-Noise Channel// *Bell Systems Technical Journal*. September 1960. Vol. 39. P. 1253–1265.
- [160] Harrison P.G., Coury S. Waiting time distribution in a class of wireless multi-channel local area networks// *Proc. Int. Conf. on Local and Metropolitan Communication Systems. Kyoto, Japan, December 7-9. 1994*. P. 409–428.
- [161] *Higher-Speed Physical Layer Extension in the 2.4 GHz Band*. Supplement to [222].
- [162] Ho T.S., Chen K.C. Performance Analysis of IEEE 802.11 CSMA/CA Medium Access Control Protocol// *Proceedings of PIMRC'96*. October 1996. P. 407–411.
- [163] *IBC's DECT'99 World Congress*.
- [164] IEEE Std IEEE 802.16-2004 (Revision of IEEE Std IEEE 802.16-2001). IEEE Standard for Local and metropolitan area networks. Part 16: Air Interface for Fixed Broadband Wireless Access Systems// *IEEE*. 1 October 2004.

- [165] Kardo-Sysoev A.F. et al. Ultra Wide Band Solid State Pulsed Antenna Array// «EUROEM 200» *Euro Electromagnetics*. Edinburgh. 30 May–2 June 2000.
- [166] Kardo-Sysoev A.F., Brylevsky V.I. et al. Powerful Sources of Ultrawide Band Pulsed Coherent Signals// «EUROEM 200» *Euro Electromagnetics*. Edinburgh. 30 May–2 June 2000.
- [167] Kopsel A., Ebert J.-P., Wolisz A. A Performance Comparison of Point and Distributed Coordination Function of an IEEE 802.11 WLAN in the Presence of Real-Time Requirements// *Proc. of 7th Int. Workshop MoMuc2000–Waseda*. Japan. October 2000.
- [168] Kuhn W. B. *Design of Integrated, Low Power, Radio Receivers in BiCMOS Technologies*: Dissertation of Doctor of Philosophy, Blacksburg, Virginia Polytechnic Institute and State University. December 1995. P. 295.
- [169] Leland W.E., Taggu M.S., Willinger W., Wilson D.V. On the self-similar nature of Ethernet traffic (extended version)// *IEEE ACM Trans. on Networking*. 1994. Vol. 2. No. 1. P. 1–15.
- [170] Lindheimer C., Mazur S., Molno J., Waleij M. Third-generation TDMA// *Ericsson Review*. 2000. No. 2.
- [171] Lindstrand Balloons Ltd.// <http://www.lindstrand.co.uk>
- [172] Low Cost Very High IP3 Low Noise Amplifiers Cover 250MHz to 2300MHz// *Mini-Circuit's*// www.minicircuits.com/
- [173] Lyakhov A.I., Vishnevsky V.M. Optional Tools of the Wi-Fi protocol: Study in Saturation// *Proc. Int. Workshop «Distributed Computer and Communication Networks (Stochastic Modelling and Optimization)» (DCCN-2003)*. Moscow, Russia, 2003. P. 28–44.
- [174] Lyakhov A.I., Vishnevsky V.M. Packet Fragmentation in Wi-Fi Ad Hoc Networks with Correlated Channel Failures// *Proc. 1st IEEE Int. Conf. on Mobile Ad-hoc and Sensor Systems (MASS 2004)*. October 24–27, 2004. Fort Lauderdale, Florida, USA. P. 204–213.
- [175] Mandelbrot B.B. *The Fractal Geometry of Nature*. San Francisco: Freeman, 1983.
- [176] Mangold S., Choi S., May P., Klein O., Hiertz G., Stibor L. IEEE 802.11e Wireless LAN for Quality of Service// *Proc. European Wireless'02*. Florence, Italy, February 2002.
- [177] Mayor T. Wi-Fi hot spots heat up// *Electronic Business*, 5/1/2000.
- [178] Nilson M. Third-generation radio access standards// *Ericsson Review*. 1999. No. 3.
- [179] Notice of Inquiry// *FCC*, 1 September, 1998.
- [180] Notice of Proposed Rule Making// *FCC*, 10 May, 2000.
- [181] Nyquist H. *Phys. Rev.* 1928. Vol. 32. P. 110.
- [182] O'Hara B., Petrick A.I. *The IEEE 802.11 Handbook: A Designer's Companion*. Inst Elect & Electronic Engineers, 1999.

- [183] Olexa R. *Implementing 802.11, 802.16, and 802.20 Wireless Networks: Planning, Troubleshooting, and Operations*. Newnes, 2004.
- [184] Platforms Wireless International// <http://www.plfm.com>
- [185] Portnoy S. Towards a Knowledge-Based Economy. *Regional Assessment Report*. UNECE Materials// www.unece.org
- [186] Poupyrev P., Kosuga M., Davis P. Analysis of Wireless Message Broadcast in Large Ad Hoc Networks of PDAs// *In proceedings of Forth IEEE conference on Mobile and Wireless Communications Networks*. 2002. P. 299–303.
- [187] Puente-Baliarda C., Romeu J., Pous R., Cardama A. On the Behavior of the Sierpinski Multiband Fractal Antenna// *IEEE Transactions on Antennas and Propagation*. April 1998. Vol. 46. No. 4. P. 517–524.
- [188] *PulsON Technology. Time Modulated Ultra-Wideband for Wireless Applications*. Time Domain Corporation, 2000.
- [189] Qiao D., Choi S., Soomoro A., Shin K.G. Energy-Efficient PCF Operation of IEEE 802.11a Wireless LAN// *Proc. of INFOCOM 2002*. New York, June 2002.
- [190] Rehmet M.A., Kroplin B.H., Epperlein F., Kornmann R., Schubert R. Recent developments on high altitude platforms// *Airship Convention*. July 2000.
- [191] *Revision of part 15 of the commission's rules regarding ultra-wideband transmission systems. First report and order. FCC 02-48*. Federal Communications Commission, 2002.
- [192] Ross G.F. A Time Domain Criterion for the Design of Windband Radiating Elements// *IEEE Trans. Antennas Propagat.* 1968. Vol. 16. No. 3.
- [193] Saunders S., Heywood P., Dornan A., Bruno L., Allen L. Wireless IP: Ready or Not, Here it Comes// *Data Communications*. 1999. No. 9. P. 42–68.
- [194] Scholtz R.A. Multiple Access with Time- Hopping Impulse Modulation (invited paper)// *IEEE MILCOM'93*. 1993.
- [195] Schriber T.J. *Simulation using GPSS*. John Wiley & Sons, 1974.
- [196] Schulteis S., Waldschmidt C., Fuegen T., Wiesbeck W. Design of a Integrable Dual Band Planar Inverted F Antenna for Small Devices// <http://inet.uni2.dk/~pballing/cost284/gothenburg/schulteis.pdf>
- [197] Serra A. Antenne Per Terminali Mobili. *Universita Degli Studi Di Pisa*// http://www2.ing.unipi.it/~o1443499/Files_Appunti/Antenne_per_terminal_i_mobi-li.pdf
- [198] Shumaker Jon. High Power GaAs FET Amplifiers: Push-Pull versus Balanced Configurations// *Wireless Symposium 2001*, 12–16 February 2001, San Jose, California. <http://www.fcsi.fujitsu.com/products/MWpdf/an014.pdf>
- [199] Shurchkov I. Towards a Knowledge-Based Economy. Russian Federation. Country Assessment Report// *UNECE Materials*// www.unece.org
- [200] SkyLINC// <http://www.skylinc.co.uk/index2.php>
- [201] SkyStation// <http://www.skystation.com>

- [202] Smith C., Meyer J. *3G Wireless with 802.16 and 802.11*. McGraw-Hill Professional, 2004.
- [203] Super Capacity Solution. The World's First GSM Adaptive Array Solution// www.airnetcom.com
- [204] The HeliNet project, <http://www.helinet.polito.it>.
- [205] The Path to 4G Mobile// *Communications Week International*. Issue 260. 5 March 2001. P. 16–17.
- [206] The Temporal and Spectral Characteristics of Ultrawideband Signals// *NTIA Report 01-383*. January 2001.
- [207] Tozer T. and Grace D. High-Altitude platforms for wireless communication// *Electronics and Communications Engineering Journal*. June 2001. P. 127–137.
- [208] Ungerboeck G. Channel coding with multilevel phase signals// *IEEE Trans. Inform Theory*. July, 1999. Vol. 45. P. 1456–1467.
- [209] Vishnevsky V.M., Lyakhov A.I. 802.11 LANs: Saturation Throughput in the Presence of Noise// *Proceedings of the 2nd International IFIP TC6 Networking Conference (Networking'2002)*. Pisa, Italy. *Lecture Notes in Computer Science*. Springer-Verlag, 2002. Vol. 2345. P. 1008–1019.
- [210] Vishnevsky V.M., Lyakhov A.I. Adaptive Features of IEEE 802.11 Protocol: Utilization, Tuning and Modifications// *Proc. of 8th HP-OVUA Conf.* Berlin, June 2001.
- [211] Vishnevsky V.M., Lyakhov A.I. Estimation of Maximal TCP/IP Traffic Rate over 802.11 Network with Hidden Stations// *Proc. of Int. Seminar «Applied stochastic models and information processes»*. Petrozavodsk, September 8–13, 2002. P. 156–158. (See also: *Information Processes*. 2002. Vol. 2. No. 2, P. 270–272.)
- [212] Vishnevsky V.M., Lyakhov A.I. IEEE 802.11 Wireless LAN: Saturation Throughput Analysis with Seizing Effect Consideration// *Cluster Computing*. 2002. No. 5. P. 133–144.
- [213] Vishnevsky V.M., Lyakhov A.I., Bakanov A.S. Method for performance evaluation of wireless networks with centralized control// *Proc. Int. Conf. «Distributed Computer Communication Networks (Theory and applications)» (DCCN'99)*. Tel-Aviv, Israel, November 9–13, 1999. P. 189–194.
- [214] Vishnevsky V.M., Lyakhov A.I., Guzakov N.N. An Adaptive Polling Strategy for IEEE 802.11 PCF// *Proc. 7th Int. Symp. on Wireless Personal Multimedia Communications (WPMC'04)*. Abano Terme, Italy, September 12–15, 2004. Vol. 1. P. 87–91.
- [215] Vishnevsky V.M., Tereshchenko B.N., Lyakhov A.I. «Wireless Communication Networks Using HAPs on the Base of Tethered Balloons» (invited paper)// *Proc. 6th Int. Symp. on Wireless Personal Multimedia Communications (WPMC'03)*. Yokosuka, Kanagawa, Japan, October 19–22, 2003. Vol. 2. P. 463–467.

- [216] Wang H.S., Moayeri N. A Useful Model for Radio Communication Channel// *IEEE Transactions on Vehicular Technology*. February 1995. Vol. 44. P. 163–171.
- [217] Warren J., Sargologos N. *PRISM NITRO Introduction*. Intersil, 2003.
- [218] Webb W. ULTRA WIDEBAND: an Electronic Free Lunch? *EDN*, 21 December, 2000.
- [219] Weinmiller J., Schlager M., Festag A., Wolisz A. Performance Study of Access Control in Wireless LANs — IEEE 802.11 DFWMAC and ETSI RES 10 HIPERLAN// *Mobile Networks and Applications*. 2(1) (1997). P. 55–76.
- [220] *Wideband wireless digital communications*/ Ed. Andreas F. Molisch. Prentice Hall PTR, 2001.
- [221] Winters J. H. Smart Antennas for Third Generation TDMA (EDGE)// *AT&T Labs-Research*, Middletown, November 27, 2000// <http://www.sce.carleton.ca/bcws/oldsource/softlibrary/27Nov00.pdf>
- [222] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*. ANSI/IEEE Std 802.11, 1999 Ed.
- [223] WO Patent № 01 82410 A1. H01Q 1/32, 1/36, 5/00. Multilevel Advanced Antenna for Motor Vehicles. Nov. 1, 2001.
- [224] WO Patent № 99/57784 A1. H01Q 5 00, H04Q 7/30. Dual Multitriangular Antennas for GSM and DCS Cellular Telephony. Nov. 11, 1998.
- [225] Wong Kin-Lu. *Planar Antennas for Wireless Communications*. N. Y.: Wiley-Interscience, 2003.
- [226] www.metawave.com
- [227] www.wireless.ru/wireless/wrl_analisis5
- [228] www.sotovik.ru/library/statistika_world.htm
- [229] www.sotovik.ru/news_CDMA.htm
- [230] www.sotovik.ru/ratings/market20010101.htm
- [231] www.wireless.ru/wireless/wrl_analisis5
- [232] Yang X. U.S. Patent № 6300914B1. H01Q 11/12// *Fractal Loop Antenna*. Oct. 9, 2001.
- [233] Yang X., Chiochetti J., Papadopoulos D., Susman L. Fractal Antenna Elements and Arrays. *Applied Microwave & Wireless*. May 1998. P. 34–46// <http://www.amwireless.com/archives/1999/May1999/may1999-p34.pdf>
- [234] Yankee Group о мировом рынке широкополосного доступа// *Телеком форум*. Вести. 2004. июль// http://www.telecomforum.ru/vesti/2004/07/14_03.htm
- [235] Ziouva E., Antonakopoulos T. The IEEE 802.11 distributed coordination function in small-scale ad-hoc wireless LANs// *International Journal of Wireless Information Networks*. 2003. Vol. 10. No. 1.
- [236] Zorzi M., Rao R.R. On the Statistics of the Block Errors in bursty channels// *IEEE Trans. On Communications*. 1997. Vol. 45. No. 6. P. 660–667.

Заявки на книги присылайте по адресу:
125319 Москва, а/я 594
Издательство «Техносфера»
e-mail: knigi@technosfera.ru
факс: (095) 956 33 46

В заявке обязательно указывайте
свой почтовый адрес!

Подробная информация о книгах на сайте
<http://www.technosfera.ru>

**В. М. Вишневский, А. И. Ляхов,
С. Л. Портной, И. В. Шахнович**
**Широкополосные беспроводные сети
передачи информации**

Компьютерная верстка — С. А. Кулешов
Корректор — О. Ч. Кохановская
Ответственный за выпуск — Л. Ф. Соловейчик

Формат 70×100/16. Печать офсетная.
Гарнитура Computer modern LaTeX.
Печ. л. 37. Тираж 2500 экз. Заказ № 900.
Бумага офсет. № 1, плотность 65г/м².

Издательство «Техносфера»
Москва, Лубянский проезд, дом 27/1

Отпечатано с готовых диапозитивов в типографии
ФГУП «Издательство «Самарский Дом печати»
443080, г. Самара, пр. К. Маркса, 201.
Качество печати соответствует качеству предоставленных диапозитивов.

ISBN 5-94836-049-0



9 785948 360492 >