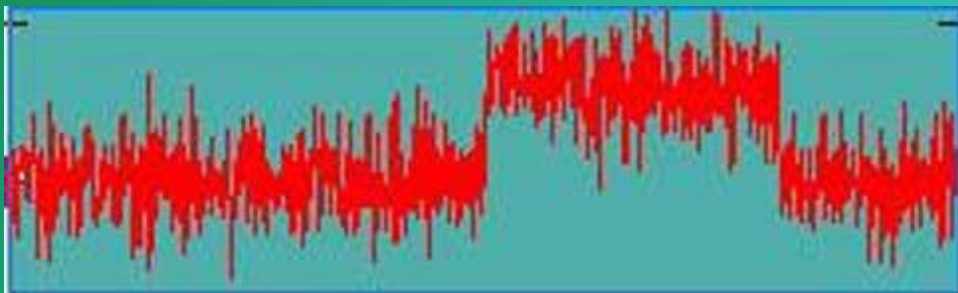


Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования

Новгородский государственный университет
имени Ярослава Мудрого

Рассветалов Л.А.

СИСТЕМЫ ЦИФРОВОЙ СВЯЗИ



Великий Новгород
2011 г.

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования

Новгородский государственный университет
имени Ярослава Мудрого

Рассветалов Л.А.

СИСТЕМЫ ЦИФРОВОЙ СВЯЗИ

УЧЕБНОЕ ПОСОБИЕ

Великий Новгород
2011 г.

ББК

УДК 621.396; 681.3.07

Рассветалов Л.А. Системы цифровой связи: Учебное пособие /
ФГБОУ «Новгородский государственный университет им. Ярослава
Мудрого», Великий Новгород, 2011 г. - _____ с.

Рецензенты: д-р техн. наук, проф. Голик Ф.В.

В учебном пособии рассмотрены вопросы: представление информации в цифровом виде, каналы передачи цифровой информации, оптимальные решающие правила приема двоичных и многоосновных сигналов, конструирование ансамблей сигналов, широкополосные методы борьбы с многолучевостью, сотовые системы мобильной связи, технологии беспроводной передачи интернет-данных.

Учебное пособие отвечает новым образовательным стандартам и предназначено для подготовки магистров по направлению 210400.68 «Системы и устройства передачи, приема и обработки сигналов».

Учебное пособие одобрено советом института Электронных и Информационных Систем Новгородского государственного университета имени Ярослава Мудрого.

© Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
Новгородский государственный университет
имени Ярослава Мудрого, 2011

Содержание

Предисловие	6
ВВЕДЕНИЕ	7
1 Структурная схема ЦСС и преобразования сигналов	9
1.1.1 Основная терминология цифровой связи	12
1.1.2 Пример сообщений, знаков и символов	13
1.1.3 Цифровые и аналоговые критерии производительности	14
1.2 Основные характеристики ЦСС	15
1.2.1 Предельные возможности ЦСС. Выбор вида сигналов	15
1.2.2 Выбор основания системы счисления	18
1.2.3 Основные коды	18
1.3 Модели каналов	21
1.3.1 Обобщенная схема ЦСПИ	21
1.3.2 Модели непрерывных каналов	22
1.3.3 Модели дискретных каналов	24
2 Задачи приема и синтез сигналов	28
2.1 Общая задача приема, оптимальные решающие правила	28
2.2 Передача двоичных данных (детерминированные сигналы)	32
2.2.1 Прием сигналов БФМ	36
2.2.2 Прием сигналов ОФМ	38
2.3 Передача М-ичных данных	41
2.4 М-ичная передача данных. Некогерентные сигналы	48
2.4.1 Некогерентные двоичные сигналы	51
2.4.2 Сравнение когерентного и некогерентного способов приема бинарных сигналов 54	
2.5 Примеры множеств ортогональных сигналов	55
2.6 Обмен между выигрышем от ортогонального кодирования и шириной полосы	61
3 Каналы со случайными параметрами	65
3.1 Вероятность ошибки приема флуктуирующих сигналов	69
3.1.1 Когерентный прием двоичных сигналов	70
3.1.2 Некогерентный прием двоичных сигналов	70
3.2 Разнесение сигналов	71
3.2.1 Широкополосные методы борьбы с многолучевостью	72
3.2.2 Установка пакетной радиосвязи (УПР)	89
4 Системы мобильной связи	92
4.1 Понятие о модели взаимодействия открытых систем	92
4.2 Структура сетей мобильной связи	94
4.3 Классификация и эволюция систем мобильной радиосвязи	95
4.3.1 Основные типы систем мобильной связи	95
4.3.2 Транкинговые системы связи	96
4.3.3 Системы персонального радиовызова	98
4.3.4 Сотовые системы мобильной связи	99
4.3.5 Эволюция систем и стандартов сотовой связи	101
4.3.6 Описание CCMC на примере GSM	102
4.3.7 Структура TDMA кадров и формирование сигналов в стандарте GSM	106
4.4 Критерии выбора модуляционных форматов при цифровой передаче данных	109
4.5 Модуляционные форматы цифровых стандартов сотовой связи второго и третьего поколений	111

4.6	Радиоинтерфейс системы cdmaOne (IS-95).....	116
4.6.1	Общая характеристика системы CDMAOne	116
4.6.2	Архитектура линии "вниз"	119
4.6.3	Пилотный канал	120
4.6.4	Канал синхронизации	121
4.6.5	Регулировка мощности в стандарте IS-95	124
4.7	Общая характеристика 3G – 4G поколений ССМС.....	125
4.7.1	Основные стандарты 3G.....	126
4.7.2	Промежуточный стандарт 3,5 G.....	129
4.7.3	Широкополосные технологии доступа к интернет-данным.....	130

Предисловие

Обилие учебников и монографий, посвященных передаче цифровой информации радиотехническими методами, не стимулирует, казалось бы, написание учебного пособия по столь популярной теме, тем более что оно носит компилятивный характер. Главным побудительным мотивом составителя пособия явилось появление нового образовательного стандарта специальности «Радиотехника», в соответствии с которым в Новгородском университете курсу «Радиотехнические системы передачи информации» предшествуют дисциплины «Статистическая радиотехника», «Теория информации и кодирование» и «Радиотехнические системы». Существующие на сегодняшний день книги учебной направленности по цифровым системам передачи информации содержат, как правило, разделы цитированных дисциплин наряду с информацией более общего характера. В результате такие методически прекрасно написанные книги, как [1, 3], имеют весьма большой объем, который зачастую просто отпугивает студентов. Поэтому составитель взял на себя смелость выбора из различных источников отдельных фрагментов и написания связывающую их часть текста, предполагая получить компактный объем настоящего пособия. Нет нужды подчеркивать, что выбор источников отражает личные пристрастия составителя

Пособие состоит из четырех глав. Первая глава содержит общие сведения о структуре цифровой системы связи (ЦСС), используемых сигналах и кодах, моделях непрерывных и дискретных каналов. В этой главе использованы материалы популярных учебников [1, 3, 4].

Во второй главе рассматривается общая задача приема сообщений. Используется критерий минимума расстояния между сигналами для формулировки оптимальных решающих правил приема сигналов с полностью известными параметрами в детерминированных каналах. Эта часть главы практически без изменений заимствована из великолепно написанной книги В. Ипатова [2]. Описание некогерентных методов приема бинарных и многоосновных сигналов базируется на материалах [2, 4].

Третья глава посвящена приему сигналов в каналах со случайными параметрами. При описании общих вопросов использованы фрагменты книг [2, 4]. Широкополосные методы борьбы с многолучевым распространением радиоволн, структуры квазиоптимальных приемников и их помехоустойчивость описываются по материалам обзора [7]. В конце этой главы помещено краткое описание установки пакетной радиосвязи [8], в которой применен остроумный способ создания опорного напряжения для детектирования широкополосных сигналов с абсолютной фазовой манипуляцией в многолучевом канале подвижной радиосвязи.

Четвертая глава «Системы мобильной связи» написана по материалам [2, 5 и 9], а также многочисленных статей на соответствующих сайтах Интернета.

Л. Рассветалов

Современная цивилизация немыслима без использования разнообразных средств передачи информации. По мере технического и социального прогресса требования к объемам передаваемых данных, качеству и широте охвата информационного сервиса непрерывно возрастают, стимулируя исследовательскую и инженерную активность и рекрутируя в эту быстро развивающуюся область все новые и новые людские резервы.

Надежность передачи сообщений в реальных линиях связи критически зависит от помех и искажений в канале, так что проблема помехоустойчивости оказывается центральной при разработке большинства информационных систем. Детерминистский подход, отвечавший концепциям начального исторического этапа теории и техники связи, с течением времени уступил позиции обобщающим и плодотворным вероятностным воззрениям, позволяющим выявлять и наилучшим образом использовать потенциальные резервы помехоустойчивой передачи информации. Современная теория связи, базирующаяся на статистических методах, включает множество ветвей и направлений, среди которых, помимо теории информации и кодирования важнейшее место занимает теория оптимального приема.

Содержание теории оптимального приема сообщений составляют статистический синтез приемных устройств, анализ эффективности различных видов модуляции, расчет показателей помехоустойчивости линии связи и др. Теория информации занимается статистическим описанием источников сообщений и каналов связи, методами экономного представления информации, а также выяснением предельных возможностей каналов с учетом энергетических и спектральных ограничений. Важным разделом теории информации, выделившимся в самостоятельную дисциплину, является теория кодирования, основной задачей которой является разработка методов отображения сообщений в некоторые абстрактные эквиваленты с целью повышения достоверности передачи данных по зашумленным каналам.

Оставляя намеренно в стороне выдающиеся физические открытия, приведшие к возникновению и повсеместному внедрению беспроводной связи, укажем лишь исторические события, знаменовавшие значительные сдвиги в понимании теоретико-информационных идей.

- Изобретение электрического телеграфа и азбуки Морзе (1837г.). Самюэль Морзе (S.Morse) предложил двоичный код переменной длины, в котором буквы английского алфавита представлены последовательностью точек и тире (кодовые слова). В этом коде часто встречающиеся буквы представлены короткими кодовыми словами, тогда как появляющиеся редко – длинными. Таким образом, код Морзе предвосхитил методы экономного неравномерного кодирования.

- Изобретение буквопечатающего телеграфа и азбуки Бодо (1875 г.). Эмиль Бодо (E.Baudot) изобрел код для телеграфной связи, в котором каждая буква кодировалась двоичным кодом фиксированной длины 5. В коде Бодо элементы двоичного кода имеют равную длину и именуются посылкой и паузой.

- Появление работ Найквиста (H.Nyquist) и Хартли (R.Hartley), посвященных исследованию канала связи и скорости передачи информации (1924, 1928 гг.) и положивших начало современной теории цифровой связи. Найквист пытался определить оптимальную форму импульса с ограниченной полосой и максимизировать скорость передачи данных в предположении, что импульс данной формы не вызывает межсимвольной интерференции. Хартли рассмотрел вопрос о количестве данных, которые могут быть переданы надежно по каналу с ограниченной полосой частот, когда для последовательной передачи данных используются импульсы со многими амплитудными уровнями.

- Установление математических основ передачи информации по каналам связи и фундаментальных ограничений для систем цифровой связи в работах Шеннона (C.Shannon) в 1947-48 гг. Шеннон сформулировал основную проблему надежной передачи информации в статистических терминах, используя вероятностные модели источников и каналов связи, и предложил универсальную логарифмическую меру количества информации. Шеннон также установил существование для любого канала основополагающей константы – пропускной

способности, т.е. предельной скорости безошибочной передачи, зависящей от физических ресурсов, доступных передатчику, и интенсивности помех в канале. Фундаментальные открытия Шеннона, часть из которых фигурирует в литературе под названием теорем Шеннона, составили базис теории информации, обязанной Шеннону также и своим наименованием.

- Открытие первых корректирующих кодов Хэммингом (R.Hamming) и Голеем (M.Golay) в 1949-50 гг. В работах названных авторов впервые были разработаны методы кодирования сообщений (т.е. отображения их в последовательности некоторых символов), позволяющие обнаруживать и исправлять ошибки, вызванные канальными помехами.

- Опубликование в 1952 г. работы Хаффмена (D.Huffman), в которой предложен высокоэффективный алгоритм сжатия данных источника сообщений.

- Выход в 1959-60 гг. работ Боуза, Рой-Чоудхури и Хоквингема (R.Bose, D.Ray-Chaudhuri, A.Nochquenghem), введших в обращение БЧХ-коды.

- Разработка в 1967 г. алгоритма декодирования Витерби (A.Viterbi) сверточных кодов, открытых Элиасом (P.Elias) в 1955 г.

- Разработка в 1976-78 гг. основ криптографии с открытым ключом в работах Диффи, Хеллмана, Ривеста, Шамира, Адлемана и др. (W.Diffie, M.Hellman, R.Rivest, A.Shamir, A.Adleman and others).

- Разработка в 1977-78 гг. алгоритма компрессии данных, описанного в работах Лемпеля (A.Lempel) и Зива (J.Ziv).

- Открытие в 1993 г. турбо-кодов, описание которых представлено в работах C.Berrou, A.Glavieux, P.Titimajshima.

В настоящем курсе излагаются идеи и технологии, являющиеся фундаментальными для систем цифровой связи. Последние становятся (и стали) все более привлекательными из-за того, что цифровая передача предлагает возможности обработки информации, недоступные при использовании аналоговой передачи.

Отличительной особенностью систем цифровой связи – СЦС (DCS – digital communication system) является передача конечного набора элементарных сигналов за конечный промежуток времени. Задачей приемника при этом является не точное воспроизведение переданного сигнала, а определение, какой именно сигнал из конечного набора был передан.

В аналоговых системах сигнал состоит из бесконечного множества элементарных сигналов, что вызывает значительные трудности достоверного приема при наличии помех.

Основным преимуществом цифрового подхода является легкость восстановления пораженных помехой цифровых сигналов по сравнению с аналоговыми. Так, при двоичном представлении информации помеха должна быть достаточно мощной, чтобы трансформировать сигнал из одного состояния в другое.

Цифровые каналы менее подвержены искажению и интерференции, т.к. наличие всего двух состояний облегчает восстановление сигнала и предотвращает накопление шумов в процессе передачи. Применение процедур выявления и коррекции ошибок делают возможным высокую точность передачи сообщения. С аналоговыми технологиями подобные процедуры недоступны.

Существуют и другие важные преимущества цифровой связи. Цифровые каналы надежнее и имеют более низкую стоимость по сравнению с аналоговыми. Цифровое программное обеспечение допускает более гибкую реализацию, чем аналоговое (микропроцессоры, цифровые коммутаторы и большие интегральные схемы – БИС). Уплотнение каналов с временным разделением – ВРК (time-division multiplexing – TDM) проще уплотнения с частотным разделением – ЧРК (frequency-division multiplexing – FDM) в аналоговых системах. При передаче различные типы цифровых сигналов (данные, телеграф, телефон, телевидение) могут рассматриваться как идентичные – бит есть бит.

1 Структурная схема ЦСС и преобразования сигналов

Функциональная схема, приведенная на рис. 1.1, иллюстрирует распространение сигнала и этапы его обработки в типичной системе цифровой связи (ЦСС).

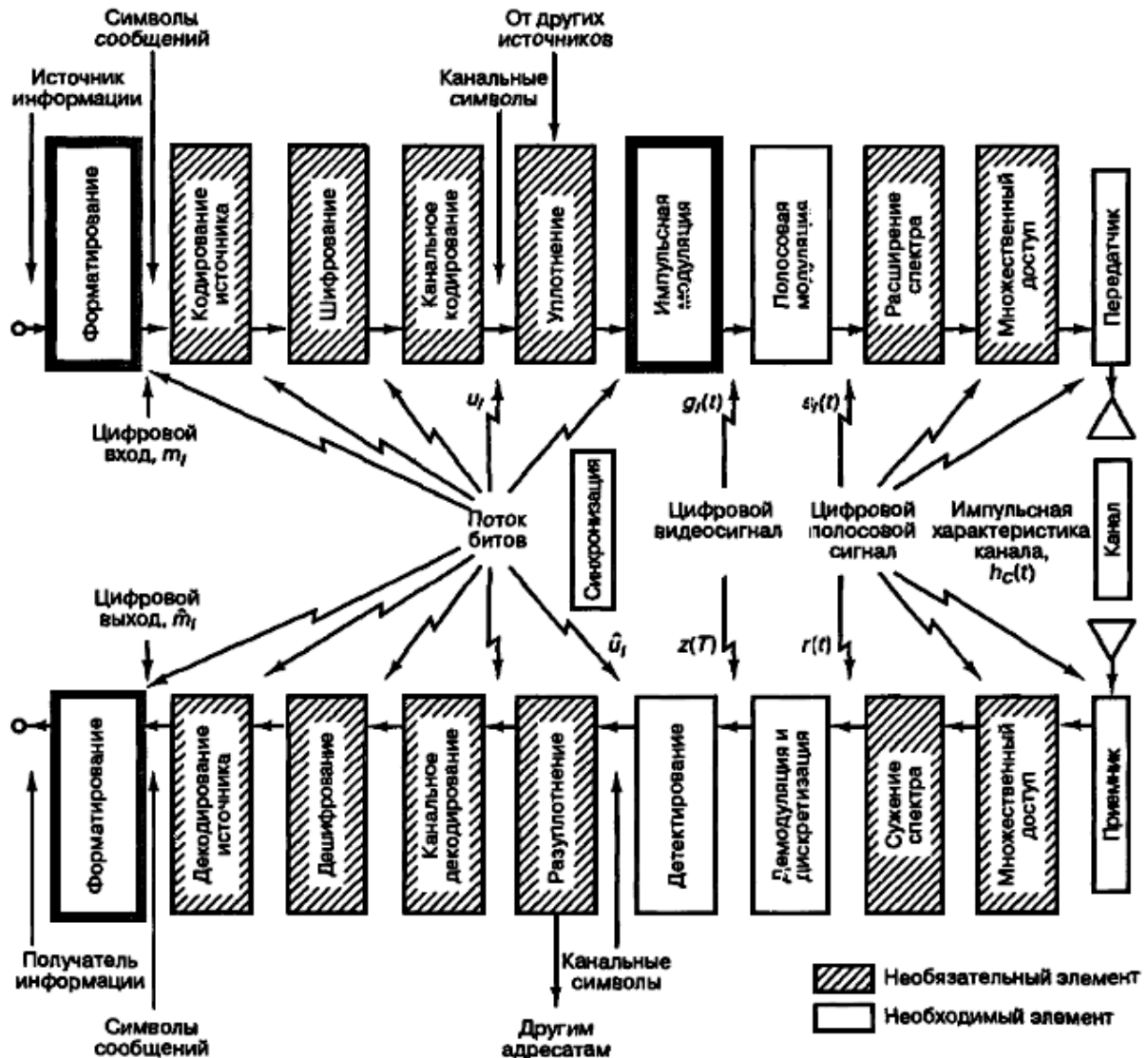


Рис. 1.1. Общая структурная схема цифровой связи.

Верхние блоки – форматирование, кодирование источника, шифрование, канальное кодирование, уплотнение, импульсная модуляция, полосовая модуляция, расширение спектра и множественный доступ — отражают преобразования сигнала на пути от источника к передатчику. Нижние блоки— преобразования сигнала на пути от приемника к получателю информации, и, по сути, они противоположны верхним блокам, блоки *модуляции* и *демодуляции/детектирования* вместе называются *модемом*. Термин "модем" часто объединяет несколько этапов обработки сигналов, показанных на рис. 1.1; в этом случае модем можно представлять как «мозг» системы. Передатчик и приемник можно рассматривать как "мускулы" системы. Для беспроводных приложений передатчик состоит из схемы повышения частоты в область радиочастот, усилителя мощности и антенны, а приемник – из антенны и малошумящего усилителя. Обратное понижение частоты производится на выходе приемника и/или демодулятора.

На рис. 1.1 исходная информация преобразуется в двоичные цифры (*биты*); после этого биты группируются в *цифровые сообщения* или *символы сообщений*. Каждый такой символ (m_i , где $i = 1, \dots, M$) можно рассматривать как элемент конечного алфавита, содержащего M элементов. Следовательно, для $M = 2$ символ сообщения m_i является бинарным (т.е. состоит из одного бита).

Рассмотрим ключевые блоки обработки сигналов, изображенные на рис. 1.1. Необходимыми для систем цифровой связи являются только этапы форматирования, модуляции, демодуляции/детектирования и синхронизации.

Форматирование преобразовывает исходную информацию в биты, обеспечивая, таким образом, совместимость информации и функций обработки сигналов с системой ЦСС. С этой точки рисунка и вплоть до блока импульсной модуляции информация остается в форме *потока битов*.

Модуляция — это процесс, посредством которого символы сообщений или каналные символы (если используется каналное кодирование) преобразуются в *сигналы*, совместимые с требованиями, налагаемыми каналом передачи данных. Канальное *не помехоустойчивое* кодирование используется в многоканальных системах, передающих сообщения от нескольких источников одновременно. В этом случае цель канального кодирования – придание сигналам различных каналов свойств, которые позволяют на приемной стороне выделить нужный канал из *группового* сигнала. Обычно канальное кодирование осуществляется с помощью *канальных поднесущих*, отличающихся частотой или формой колебаний. Подвергнувшиеся канальному кодированию сигналы разных источников поступают в блок *уплотнения*, где формируется групповой сигнал. *Помехоустойчивое* кодирование может осуществляться как при канальном кодировании, так и над групповым сигналом. Специфические вопросы многоканальной передачи будут рассмотрены отдельно; дальнейшее описание предполагает наличие одного источника (канала).

Импульсная модуляция — это еще один необходимый этап, поскольку каждый символ, который требуется передать, вначале нужно преобразовать из двоичного представления в *видео-сигнал* (*модулированный сигнал*). Блок импульсно-кодовой модуляции обычно включает фильтрацию с целью достижения минимальной полосы передачи. При использовании импульсной модуляции для обработки двоичных символов результирующий двоичный сигнал называется ИКМ-сигналом (ИКМ – импульсно-кодовая модуляция, в англоязычной литературе – PCM - pulse-code modulation). Существует несколько типов ИКМ-кодированных сигналов; в приложениях телефонной связи эти сигналы часто называются *кодами канала*. При применении импульсной модуляции к не двоичным символам результирующий сигнал именуется М-арным (М-ичным) импульсно-модулированным. Существует несколько типов подобных сигналов, основное внимание будем уделять *амплитудно-импульсной модуляции* АИМ (*англ.* pulse-amplitude modulation — PAM).

После импульсной модуляции каждый символ сообщения или каналный символ принимает форму видеосигнала $g_i(t)$, где $i = 1, \dots, M$. В любой электронной реализации поток битов, предшествующий импульсной модуляции, представляется уровнями напряжений. Может возникнуть вопрос, почему существует отдельный блок для импульсной модуляции, когда фактически уровни напряжения для двоичных нулей и единиц уже можно рассматривать как идеальные прямоугольные импульсы, длительность каждого из которых равна времени передачи одного бита? Существует два важных отличия между подобными уровнями напряжения и видеосигналами, используемыми для модуляции. Во-первых, блок импульсной модуляции позволяет использовать бинарные и М-арные сигналы. Во-вторых, фильтрация, производимая в блоке импульсной модуляции, формирует импульсы, длительность которых больше времени передачи одного бита. Фильтрация позволяет использовать импульсы большей длительности; таким образом, импульсы расширяются на соседние временные интервалы передачи битов. Этот процесс иногда называется формированием импульсов; он ис-

пользуется для поддержания полосы передачи в пределах некоторой желаемой области спектра.

Для систем передачи радиочастотного диапазона следующим важным этапом является *полосовая модуляция*; она необходима всегда, когда среда передачи не поддерживает распространение видеосигналов, а требует радиочастотного сигнала $s_i(t)$, где $i = 1, \dots, M$. Далее сигнал s_i проходит через канал, причем связь между входным и выходным сигналами канала полностью определяется *импульсной характеристикой канала* $h_c(t)$. Кроме того, в различных точках вдоль маршрута передачи дополнительные случайные шумы искажают сигнал, так что сигнал на входе приемника $y(t)$ отличается от переданного сигнала $s_i(t)$:

$$y(t) = s_i(t) * h_c(t) + n(t), \quad i = 1, \dots, M, \quad (1.1)$$

где знак «*» представляет собой операцию свертки, а $n(t)$ — случайный процесс.

При обработке полученного сигнала в принимающем устройстве входной каскад приемника и/или демодулятор обеспечивают понижение частоты каждого полосового сигнала $y(t)$. В качестве подготовки к детектированию демодулятор восстанавливает $y(t)$ в виде оптимальной огибающей видеосигнала $z(t)$. Обычно с приемником и демодулятором связано несколько фильтров — фильтрация производится для удаления нежелательных высокочастотных составляющих (в процессе преобразования полосового сигнала в видеосигнал) и формирования импульса. Выравнивание можно описать как разновидность фильтрации, используемой в демодуляторе (или после демодулятора) для удаления всех эффектов ухудшения качества сигнала, причиной которых мог быть канал. Выравнивание (*англ.*: equalization) необходимо в том случае, если импульсная характеристика канала $h_c(t)$ настолько плоха, что принимаемый сигнал сильно искажен.

Эквалайзер (устройство выравнивания) реализуется для компенсации (т.е. для удаления или ослабления) всех искажений сигнала, вызванных неидеальной характеристикой $h_c(t)$. И последнее, на этапе дискретизации сформированный импульс $z(t)$ преобразовывается в выборку $z(T)$ для восстановления (приблизительно) символа канала и/или символа сообщения m_i (если не используется канальное кодирование). Некоторые авторы используют термины "демодуляция" и "детектирование" как синонимы. Здесь под *демодуляцией* подразумевается восстановление сигнала (полосового импульса), а под *детектированием* — принятие решения относительно цифрового значения этого сигнала.

Далее, однако, будем считать, что обе операции — демодуляция и детектирование выполняются в одном устройстве — демодуляторе.

Остальные этапы обработки сигнала в модеме являются необязательными и направлены на обеспечение специфических системных нужд.

Кодирование источника — это преобразование аналогового сигнала в цифровой (для аналоговых источников) и удаление избыточной (ненужной) информации, т.е. *статистическое кодирование*. Отметим, что типичная система ЦСС может использовать либо *кодирование источника* (для оцифровывания и сжатия исходной информации), либо более простое *форматирование* (только для оцифровывания). Система не может одновременно применять и кодирование источника, и форматирование, поскольку первое уже включает необходимый этап оцифровывания информации.

Шифрование, которое используется для обеспечения секретности связи, предотвращает понимание сообщения несанкционированным пользователем и введение в систему ложных сообщений.

Канальное кодирование при данной скорости передачи данных может снизить вероятность ошибки P_E (помехоустойчивое кодирование) или уменьшить отношение сигнал/шум, необходимое для получения желаемой вероятности P_E за счет увеличения полосы передачи или усложнения декодера.

Процедуры *уплотнения* и *множественного доступа* объединяют сигналы, которые могут иметь различные характеристики или могут поступать от разных источников, с тем, чтобы они могли совместно использовать часть ресурсов связи (например, спектр, время).

Расширение полосы частот может давать сигнал, относительно неуязвимый для узкополосной помехи (как естественной, так и умышленной), и может использоваться для повышения конфиденциальности сеанса связи. Также оно является ценной технологией, используемой для множественного доступа и работы в многолучевых каналах.

Блоки обработки сигналов, показанные на рис. 1.1, представляют типичную функциональную схему системы цифровой связи; впрочем, эти блоки иногда реализуются в несколько ином порядке. Например, уплотнение может происходить до канального кодирования *или* модуляции *либо* — при двухэтапном процессе модуляции (поднесущая и несущая) — оно может выполняться между двумя этапами модуляции. Подобным образом блок расширения частоты может находиться в различных местах верхнего ряда рис. 1.1; точное его местонахождение зависит от конкретной используемой технологии. Синхронизация и ее ключевой элемент, синхронизирующий сигнал, задействованы во всех этапах обработки сигнала в системе ЦСС. Для простоты блок синхронизации на рис. 1.1 показан безотносительно к чему-либо, хотя фактически он участвует в регулировании операций практически в каждом блоке, приведенном на рисунке.

1.1 Сообщения, знаки и символы

1.1.1 Основная терминология цифровой связи

Ниже приведены некоторые основные термины, часто используемые в области цифровой связи.

Источник информации. Устройство, передающее информацию посредством системы ЦСС. Источник информации может быть *аналоговым* или *дискретным*. Выход аналогового источника может иметь любое значение из непрерывного диапазона амплитуд, тогда как выход источника дискретной информации — значения из конечного множества амплитуд. Источники аналоговой информации преобразуются в источники цифровой информации посредством *дискретизации* или *квантования*. Методы дискретизации и квантования называются форматированием и кодированием источника.

Текстовое сообщение. Последовательность символов. При цифровой передаче данных сообщение представляет собой последовательность цифр или символов, принадлежащих конечному набору символов или алфавиту.

Знак. Элемент алфавита или набора символов. Знаки могут представляться последовательностью двоичных цифр. Существует несколько стандартизованных кодов, используемых для знакового кодирования, в том числе код ASCII (American Standard Code for Information Interchange — Американский стандартный код для обмена информацией), код Холлерита (Hollerith code), код Бодо (Baudot code), код Муррея (Murray code) и код (азбука) Морзе (Morse code).

Двоичная цифра — *бит* (binary digit — bit). Фундаментальная единица информации для всех цифровых систем. Термин "бит" также используется как единица объема информации. *Поток битов* (bit stream) — последовательность двоичных цифр (нулей и единиц). Поток битов часто называют *видеосигналом*, или *низкочастотным сигналом*.

Символ (цифровое сообщение). Символ — это группа из k бит, рассматриваемых как единое целое. Далее мы будем называть этот блок *символом сообщения* m_i ($i = 1, \dots, M$) из конечного набора символов или алфавита (рис. 1.2) Размер алфавита M равен 2^k , где k — число битов в символе. При *низкочастотной* передаче каждый из сигналов m_i будет представлен одним из набора видеопульсов $g_1(t), g_2(t), \dots, g_M(t)$. Иногда при передаче последовательности таких импульсов для выражения скорости передачи импульсов (скорости передачи символов) используется единица *бод* (baud), характеризующая *техническую* скорость передачи. Для типичной *полосовой* передачи каждый импульс $g_i(t)$ будет представляться одним из набора полосовых импульсных сигналов $s_1(t), s_2(t), \dots, s_M(t)$. Таким образом, для беспроводных систем символ m_i посылается путем передачи цифрового сигнала $s_i(t)$ в течение T секунд (T — длительность символа). Следующий символ посылается в течение следующего временного интервала, T . То, что набор символов, передаваемых системой ЦСС, является конечным, и

есть главным отличием этих систем от систем аналоговой связи. Приемник ЦСС должен всего лишь определить, какой из возможных M сигналов был передан; тогда как аналоговый приемник должен точно определять значение, принадлежащее непрерывному диапазону сигналов.

Цифровой сигнал. Описываемый уровнем напряжения или силы тока, сигнал (импульс — для низкочастотной передачи или синусоида — для полосовой передачи), представляющий цифровой символ. Характеристики сигнала (для импульсов — амплитуда, длительность и положение или для синусоиды — амплитуда, частота и фаза) позволяют его идентифицировать как один из символов конечного алфавита.

Скорость передачи данных. Эта величина в битах в секунду (бит/с) дается формулой $R = k/T = (1/T) \log_2 M$ (бит/с), где k бит определяют символ из $M = 2^k$ -символьного алфавита, а T — это длительность k -битового символа.

Текстовые сообщения состоят из последовательности буквенно-цифровых знаков. При цифровой передаче знаки вначале кодируются в последовательность битов, которая называется *поток битов*, или *видеосигналом*. После этого формируются группы из k бит, именуемые *символами*, причем число всех символов конечно ($M = 2^k$), а их совокупность называется *алфавитом*. Система, использующая символьный набор размера M , называется *M-арной* (*M-ичной*). Выбор величины k или M является важным первоначальным этапом проектирования любой цифровой системы связи. При $k = 1$ система является *бинарной*, размер набора символов равен $M = 2$, а модулятор использует один из двух различных сигналов для представления двоичного значения "один", а другой — для представления двоичного значения "ноль". В этом частном случае символ и бит — это одно и то же. При $k = 2$ система именуется *четверичной*, или *4-уровневой* ($M = 4$). В каждый момент формирования символа модулятор использует один из четырех возможных сигналов для представления символа. Разделение последовательности битов сообщения определяется размером алфавита M . Ниже приведен пример, который поможет лучше понять связь между следующими терминами: "сообщение", "знак", "символ", "бит" и "цифровой сигнал".

1.1.2 Пример сообщений, знаков и символов

На рис. 1.2 приведен пример разбиения потока битов, определяемого спецификацией системы для различных значений k и M . Текстовое сообщение на рисунке — это слово "ДУМАЙ!". Использование 8-битовой кодировки ANSI дает поток битов, состоящий из 48 бит. Кодировка ANSI Windows 1251 приведена в Таблице 1.

Символы 0 – 127 совпадают с кодировкой ASCII, приведенной в Приложении 1.

Таблица 1. Кодировка Windows 1251

128 Ъ	144 ђ	160 Ѡ	176 ±	192 А	208 Р	224 а	240 р
129 Ѓ	145 ѓ	161 ѡ	177 ±	193 Б	209 С	225 б	241 с
130 ,	146 ‚	162 Ѣ	178 і	194 В	210 Т	226 в	242 т
131 ҃	147 ҃	163 Ј	179 і	195 Г	211 У	227 г	243 у
132 ˆ	148 ˆ	164 Ѥ	180 ґ	196 Д	212 Ф	228 д	244 ф
133 ...	149 ˙	165 Ғ	181 μ	197 Е	213 Х	229 е	245 х
134 †	150 −	166 Ҕ	182 ¶	198 Ж	214 Ц	230 ж	246 ц
135 ‡	151 −	167 §	183 ˙	199 З	215 Ч	231 з	247 ч
136 ˙	152 ˙	168 Е	184 ё	200 И	216 Ш	232 и	248 ш
137 ‰	153 ™	169 ©	185 №	201 Ъ	217 Щ	233 ъ	249 щ
138 Љ	154 ъ	170 €	186 €	202 К	218 Ь	234 к	250 ь
139 ˘	155 ˘	171 ˘	187 ˘	203 Л	219 Ы	235 л	251 ы
140 Њ	156 њ	172 ˘	188 ј	204 М	220 Ь	236 м	252 ь
141 Ќ	157 ќ	173 ˘	189 S	205 Н	221 Э	237 н	253 э
142 Ѓ	158 ґ	174 ©	190 s	206 О	222 Ю	238 о	254 ю
143 Ў	159 ӯ	175 ı	191 ı	207 П	223 Я	239 п	255 я

На рис. 1.2, а размер набора символов, M , был выбран равным 8 (каждый символ представляет восьмеричное число). Таким образом, биты группируются по три ($k = \log_2 8$); полученные в результате 16 чисел представляют 16 готовых к передаче восьмеричных символов. Передатчик должен иметь набор из восьми сигналов $s_i(t)$, где $i = 1, \dots, 8$, сопоставляемых со

всеми возможными символами, причем передача каждого сигнала возможна в течение времени символа. В последней строке рис. 1.2, а указаны 16 сигналов, передаваемых восьмеричной системой модуляции для представления текстового сообщения "ДУМАЙ!".

Сообщение (текст): «ДУМАЙ!»

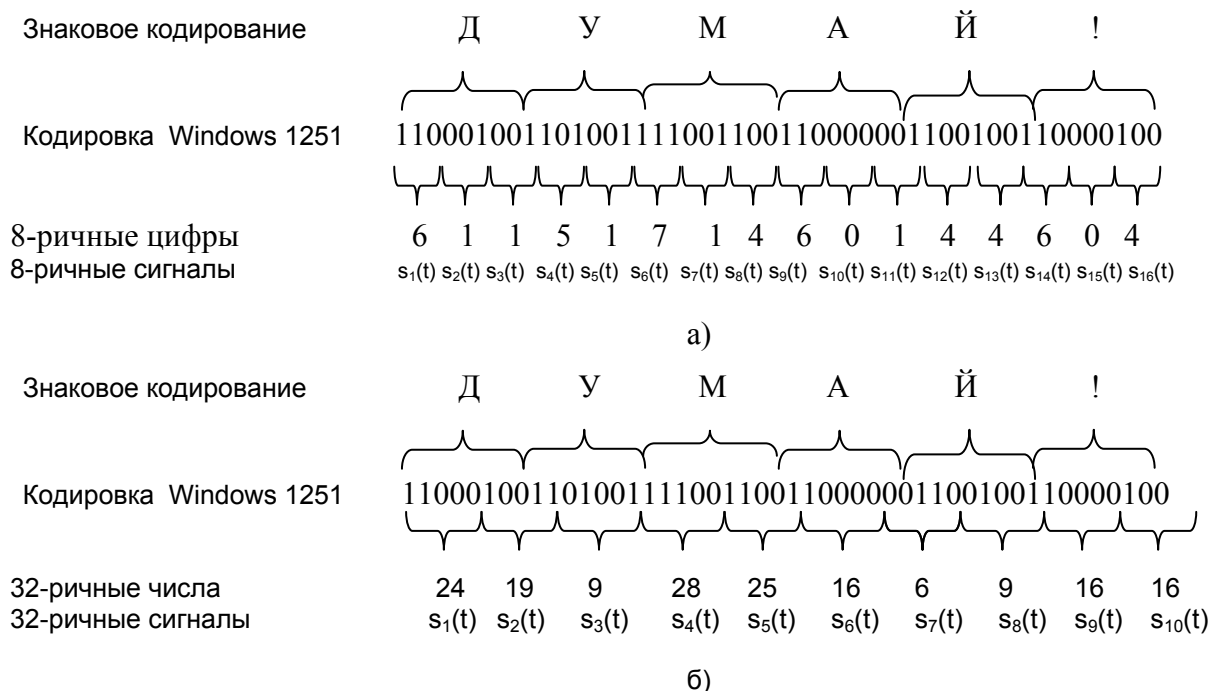


Рис. 1.2. Сообщения, знаки и символы: а) 8-ричный пример; б) 32-ричный пример.

На рис. 1.2, б размер набора символов, M , был выбран равным 32 (каждый символ представляет 32-ричную цифру). Следовательно, биты берутся по пять, а результирующая группа из десяти чисел представляет десять готовых к передаче 32-ричных символов. Отметим, что границы символов и знаков не обязательно должны совпадать. Первый символ представляет 5/8 первого знака, "Д", второй символ — оставшуюся 3/8 знака "Д" и 2/8 следующего знака, "У", и т.д. Более эффектное разбиение знаков совсем не обязательно, поскольку система рассматривает знаки как строку символов, которую необходимо передать; только конечный пользователь (или телетайп пользователя) приписывает текстовое значение полученной последовательности битов. В 32-ричном примере передатчик должен содержать набор из 32 символов $s_i(t)$, где $i=1, \dots, 32$, сопоставляемых со всеми возможными символами. На рис. 1.2, б указаны десять сигналов, передаваемых 32-ричной системой модуляции для представления текстового сообщения "ДУМАЙ!".

1.1.3 Цифровые и аналоговые критерии производительности

Принципиальное отличие систем аналоговой и цифровой связи связано со способом оценки их производительности. Сигналы аналоговых систем составляют континуум, так что приемник должен работать с бесконечным числом возможных сигналов. Критерием производительности аналоговых систем связи является критерий достоверности, такой как отношение сигнал/шум, процент искажения или ожидаемая среднеквадратическая ошибка между переданным и принятым сигналами.

В отличие от аналоговых, цифровые системы связи передают сигналы, представляющие цифры. Эти цифры формируют конечный набор или алфавит, и этот набор известен приемнику априорно. Критерием качества цифровых систем связи является вероятность неверного детектирования цифры или вероятность ошибки (P_E).

1.2 Основные характеристики ЦСС

Любая система связи характеризуется скоростью передачи R и достоверностью; последний параметр для ЦСС определяется вероятностью ошибок P_E . Однако этих характеристик недостаточно для оценки работы всей системы связи в целом. Всегда желательно, чтобы линия связи обеспечивала передачу информации с требуемым качеством и скоростью наиболее экономно, т.е. с наименьшими затратами энергетического и частотного ресурсов.

В теории связи наиболее широко применяются критерии качества работы ЦСС, в соответствии с которыми она оценивается величиной затрат на передачу единицы количества информации при заданном качестве ее приема. Такие критерии называются удельными, а под затратами понимают расход энергии, полосы частот, вес и габариты системы связи и ее стоимость. В дальнейшем будем учитывать только затраты энергии и полосы частот на передачу одной двоичной единицы передачи информации – бита.

Удельные затраты энергии будем характеризовать величиной

$$\beta_E = E_b / N_0, \quad (1.2)$$

где E_b – энергия входного сигнала приемника, соответствующая передаче одного бита информации с заданной достоверностью, N_0 – спектральная плотность мощности помехи на входе приемника.

Удельные затраты полосы частот будем оценивать величиной

$$\beta_W = W/R, \quad (1.3)$$

где W – эквивалентная (эффективная) полоса частот спектра сигнала, которой соответствует полоса пропускания приемного устройства, R – информационная скорость передачи (бит/с).

Т.о., качество работы системы цифровой связи можно характеризовать следующими показателями:

1. вероятностью ошибки при передаче одного элемента сообщения;
2. информационной и технической скоростью передачи (техническая скорость передачи R_t обратно пропорциональна длительности T_t одного элемента дискретного сообщения $R_t = 1/T_t$;
3. энергетическими затратами на передачу одного бита информации (удельные затраты энергии);
4. затратами полосы частот на передачу одного бита информации (удельные затраты полосы)

1.2.1 Предельные возможности ЦСС. Выбор вида сигналов

Для гауссовского канала с белым шумом и ограниченной средней мощностью сигнала К.Шеннон [1] получил широко известную формулу, определяющую максимум средней скорости передачи информации по такому каналу:

$$\max R = C = W \log_2 (1 + P_c / P_{ш}) = W \log_2 (1 + P_c / N_0 W), \quad (1.4)$$

где C – пропускная способность канала, P_c – средняя мощность сигнала, $P_{ш}$ – средняя мощность шума в канале, N_0 – спектральная плотность белого шума.

Эта формула характеризует предельные возможности канала, в котором реализованы оптимальные кодирование и декодирование, обеспечивающие согласование производительности источника с пропускной способностью канала. В этом случае ошибки отсутствуют. Вместе с тем заметим, что (1.4) справедлива только при физически нереализуемых условиях: бесконечном времени передачи информации и неизменных условиях работы канала¹.

Представим введенные выше затраты энергии и полосы на передачу одного бита информации в идеальном канале величинами

$$\beta_E = E_b / N_0 = P_c T_0 / N_0, \quad (1.5)$$

¹ Открытые недавно турбо-коды позволяют вплотную приблизиться к пределу Шеннона.

$$\beta_W = W/R_{max} = W/C, \quad (1.6)$$

где T_0 – время передачи одного бита информации.

Т.к. по определению $\max RT_0 = CT_0 = 1$ бит, то (1.4) с учетом (1.5) и (1.6) преобразуем к виду

$$\frac{W}{\tilde{N}} \log_2 \left(1 + \frac{E_b}{N_0} \frac{C}{W} \right) = 1 \quad (1.7)$$

или

$$\log_2 \left(1 + \frac{\beta_E}{\beta_W} \right)^{\beta_W} = 1 \quad (1.8)$$

Решая (1.8) относительно β_E , получим:

$$\beta_E = \beta_W (2^{1/\beta_W} - 1). \quad (1.9)$$

Выражение (1.9) определяет функциональную связь между удельными затратами энергии и полосы в идеальном гауссовском канале связи. Исследуя его на экстремум, получим, что при увеличении удельных затрат полосы удельные затраты энергии в идеальном канале уменьшаются, стремясь в пределе к величине

$$\beta_E \rightarrow \ln 2 \approx 0,7 \text{ при } \beta_W \rightarrow \infty. \quad (1.10)$$

Зависимость (1.10) изображена на рис.1.3. Эта зависимость называется границей Шеннона для идеального гауссовского канала связи.

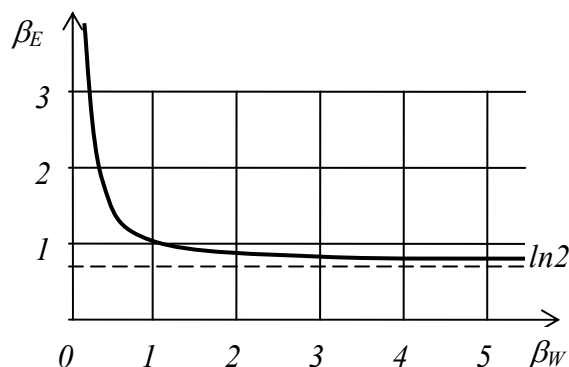


Рис. 1.3.

Она показывает, что существует бесконечное множество оптимальных систем, каждая из которых характеризуется своей парой β_E и β_W . Из графика на рис. 1.3 следует, что желание улучшить один из показателей неизбежно ухудшает другой. Оптимальные системы с малыми затратами энергии требуют значительных затрат полосы и наоборот. Понятно, что потенциально реализуемым системам соответствуют те точки на плоскости $\beta_E - \beta_W$, которые лежат выше границы Шеннона.

Проблема выбора сигнала

Из общей теории информации известно, что максимальная средняя производительность H_t источника информации определяется величиной

$$\max H_t = \lim_{T \rightarrow \infty} (\log_2 M / T), \quad (1.11)$$

где M – число различных сообщений длительностью T , вырабатываемых источником. Введем обозначения:

k – количество элементов длительностью T_0 в символе (слове),

n – алфавит источника сообщений,

$M = n^k$ – количество символов источника. Перепишем (1.11) с учетом сделанных обозначений:

$$\max R = C = 1/T_0 = \max H_t = \lim_{k \rightarrow \infty} (\log_2 n^k / kT) = \log_2 n / T, \quad (1.12)$$

которое показывает, что максимальная производительность источника в данном случае определяется только основанием алфавита n и длительностью информационного символа T .

Из (1.12) непосредственно следует

$$T_0 = T / \log_2 n. \quad (1.13)$$

Так, для кодировки ASCII, содержащей 128 символов, один информационный символ (словесное слово) содержит 7 бит, т.е. (1.13) устанавливает связь между временем T_0 на передачу одного бита и длительностью T информационного символа. Запишем выражение для удельных затрат полосы в идеальном канале с учетом (1.13):

$$\beta_W = W / \max R = W T_0 = W T / \log_2 n = B_C / \log_2 n, \quad (1.14)$$

где $B_C = W T$ – база сигнала.

Рассмотрим два характерных случая.

1) Малые затраты полосы ($\beta_W \leq 1$). В соответствии с (1.14) используемые сигналы характеризуются соотношением $B_C \leq \log_2 n$, откуда следует

$$n \geq 2^{B_C} \quad (1.15)$$

и применение сложных сигналов с большой базой потребует огромных оснований алфавита. Использование простых сигналов с базой $B_C = 1$ приводит к следующему требованию для алфавита: $\beta_W = 1 / \log_2 n$, или

$$n = 2^{1/\beta_W} \geq 2 \quad (1.16)$$

Т.о., для получения малой полосы следует выбирать простые сигналы и использовать алфавиты с основанием больше двух. Такие M -ичные сигналы часто называют *многоосновными*.

Рис. 1.3 показывает, что малые затраты полосы требуют больших затрат энергии, что иллюстрирует Таблица 2.

Таблица 2. Затраты в идеальной системе с простыми сигналами

n	2	4	8	16	32	64
β_W	1	0,5	0,33	0,25	0,2	0,167
$1/\beta_W$	1	2	3	4	5	6
β_E	1	1,5	2,31	3,75	6,2	10,5

2) Малые затраты энергии ($\beta_E \leq 1$). Это условие соответствует $\beta_W \rightarrow \infty$. Используемые при этом сигналы характеризуются соотношением

$$B_C \gg \log_2 n. \quad (1.17)$$

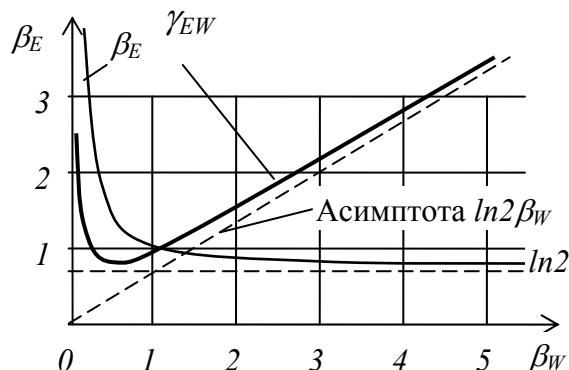
Т.к. минимум n равен двум, для малых затрат энергии необходимо применять сложные сигналы с $B_C \gg 1$.

Малые затраты полосы и энергии.

Введем коэффициент обобщенных затрат $\gamma_{EW} = \beta_E \beta_W$, который в равной степени зависит как от затрат энергии, так и полосы. Подставляя в него выражение для β_E , будем иметь

$$\gamma_{EW} = \beta_W^2 (2^{1/\beta_W} - 1) \quad (1.18)$$

Зависимость (1.18) показана на рис. 1.4 и имеет минимум при $\beta_{Wopt} = 0,435$, равный $\min \gamma \approx 0,735$.



Область $\beta_W \approx 0,35..0,5$ соответствует классу простых сигналов с основанием от 4 до 8 ($n = 4..8$, см. (1.16) и величину $1/\beta_W$ в Таблице 1). В этой области, как видно из рисунка, коэффициент γ_{EW} минимален и практически не меняется. Значит, простые сигналы с основанием кода от четырех до восьми являются оптимальными с точки зрения минимизации совместных затрат полосы частот и энергии.

Рис. 1.4.

Если рассматривать область, в которой обобщенные затраты $\gamma_{EW} \leq 1$, то ей соответствуют простые сигналы с основанием $n = 2..16$.

1.2.2 Выбор основания системы счисления

В своей повседневной деятельности мы пользуемся десятичной системой. В информационных технологиях используются системы счисления, основания которых кратно двум: двоичная, восьмеричная, шестнадцатеричная, и т. д. Возникает вопрос, какая система счисления наиболее экономно (компактно) представляет любые числа?

Если n – основание системы; m – число разрядов для записи числа, то количество элементов, используемых системой для записи числа, определяется, как $v = mn$. Максимальное число, которое может быть записано в такой системе, есть $\max N(n) = n^m - 1$.

Откуда

$$n^m = \max N(n) + 1.$$

Логарифм последнего выражения:

$$m = \ln[\max N(n) + 1] / \ln(n)$$

Умножим его на n и получим:

$$v = \frac{n}{\ln(n)} \ln[\max N(n) + 1].$$

Исследуем на экстремум:

$$\frac{dv}{dn} = \ln[\max N + 1] \frac{\ln(n) - 1}{(\ln(n))^2} = 0,$$

откуда

$$n = n_{opt} = e \approx 2,72.$$

Итак, мы получили, что наиболее компактно числовые величины представляются в системе счисления с основанием $n = e$. В Таблице 3 представлены характеристики некоторых употребительных систем счисления в сравнении с оптимальной при $n = e$.

Таблица 3. Сравнение систем счисления.

n	2	2,72	3	4	8	10	12
$n/\ln n$	2,88	2,72	2,735	2,88	3,85	4,3	4,83
v/v_{opt}	1,06	1	1,006	1,06	1,42	1,58	1,77

Видно, что системы счисления с основанием от двух до четырех практически не проигрывают оптимальной. С технологической точки зрения двоичная система несравненно удобнее остальных, что и обусловило ее широкое распространение.

1.2.3 Основные коды

Код – это система соответствий между элементами дискретных сообщений и кодовыми комбинациями. Будем рассматривать двоичные коды, используемые при форматировании символов источника. Встречаются различные обозначения символов двоичного кода, представленные таблицей 4.

Таблица 4. Обозначения символов двоичного кода

x_1	1	$+1$	$+$	Z	M	
x_2	0	-1	$-$	A	S	

Обозначения «1» и «0» как «Z» и «A» рекомендованы Международным Консультативным Комитетом по телефонии и телеграфии (МККТТ), M и S – “mark” и “space”.

Наиболее употребительные коды перечислены на рис. 1.5.

№	Вид последовательности символов								Метод формирования кода
		1	0	1	1	0	0	1	
1	0								ВН Униполярные символы с возвратом к нулю
2	0								ВН Униполярные символы с расщепленной фазой
3	0								ВН Полярные символы с расщепленной фазой
4	0								БВН Полярные символы с активной паузой
5	0								БВН Полярные символы с расщепленной фазой
6	0								ВН Относительный код <i>S</i>
7	0								ВН Относительный код <i>M</i>

Рис. 1.5. Виды наиболее употребительных кодов

Различают коды с возвратом к нулю – ВН и без возврата к нулю – БВН. В англоязычной литературе и ряде переводных книг ВН соответствует RZ (return-to-zero), а БВН – NRZ (no return-to-zero). На диаграммах 1, 2 показаны униполярные коды ВН, на диаграммах 3-5 полярные коды. Коды ВН могут принимать три значения (диаграмма 3), коды БВН – только два. Коды с расщепленной фазой передают «1» импульсом в первой половине тактового интервала (длительности бита), а «0» – во второй.

Кодировка БВН, диаграмма 4 – биполярные импульсы с активной паузой – широко используется в цифровых логических схемах и известна также под наименованием NRZ-L (L – level – уровень). Двоичная единица при этом представляется одним уровнем напряжения, а двоичный нуль – другим.

Группа кодировок с расщепленной фазой применяется в системах связи, тактовая синхронизация в которых осуществляется принимаемым сигналом – в их спектре всегда присутствует тактовая частота. Кодировка 5 известна также под названием *манчестерская* (Manchester Encoding); используется в системах магнитной записи и оптической связи, а также в некоторых спутниковых каналах телеметрической передачи данных.

Относительные коды *M* и *S* применяются в телеграфии и системах связи с относительной фазовой манипуляцией (ОФМ). Код *S* изменяет свою полярность с приходом «1», а код *M* – с приходом «0».

Код называется *равномерным*, если все кодовые комбинации содержат одинаковое количество элементов, например, код Бодо, ASCII. Код Морзе – *неравномерный*.

Код называется *простым* или *первичным*, если для передачи числа N используются m разрядов, причем $n^{m-1} < N \leq n^m$, где n – основание кода. Например, для передачи максимального числа 130, имеющегося в сообщении, *натуральным двоичным кодом* требуется 8 разрядов, тогда как такой код может кодировать 256 чисел. Значит, $256-130 = 126$ чисел не будут использованы.

Невзвешенные коды.

Натуральному двоичному коду присущ недостаток, приводящий к значительным ошибкам при действии помех. Т.к. такой код является взвешенным, т.е. разряды кодового слова имеют разный вес, ошибка в старшем разряде приводит к гораздо большему искажению сообщения, нежели ошибка в младших разрядах.

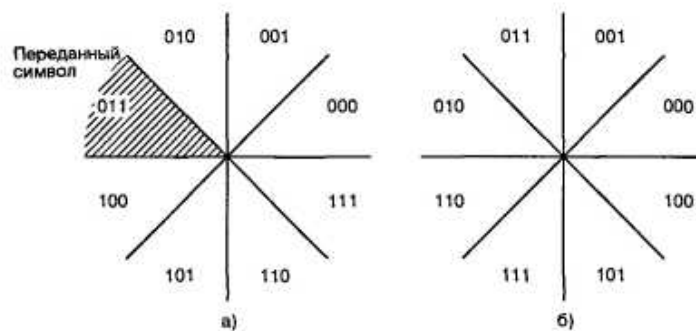


Рис.1.6. Пространства решений для: а) натурального двоичного кода; б) кода Грея.

Сказанное иллюстрируется рис. 1.6, где представлены пространства решений при использовании натурального двоичного кода (а) и невзвешенного кода Грея (б). При передаче М-ичными фазоманипулированными сигналами (МФМ) каждый сигнальный вектор не является равноудаленным от всех остальных.

На рис. 1.6, а показано восьмеричное пространство решений, где области решений обозначены 8-ричными символами в двоичной записи. При передаче символа (0 1 1) и появлении в нем ошибки наибольшую вероятность превратиться в тот же символ имеют ближайшие соседние символы, (0 1 0) и (1 0 0). Вероятность превращения символа (0 1 1) вследствие ошибки в символ (1 1 1) относительно мала. Если биты распределяются по символам согласно двоичной последовательности, показанной на рис. 1.6, а, то некоторые символьные ошибки всегда будут давать две (или более) битовые ошибки, даже при значительном отношении сигнал/шум.

Для неортогональных схем, таких как МФМ, часто используется код преобразования бинарных символов в М-арные, такой, что двоичные последовательности, соответствующие соседним символам (сдвигам фаз), отличаются только одной битовой позицией; таким образом, при появлении ошибки в М-арном символе высока вероятность того, что ошибочным является только один из k прибывших битов. Кодом, обеспечивающим подобное свойство, является код Грея; на рис. 1.6, б для восьмеричной схемы ФМ показано распределение битов по символам с использованием кода Грея. Можно видеть, что соседние символы отличаются одним двоичным разрядом. Следовательно, вероятность появления многобитовой ошибки при данной символьной ошибке значительно меньше по сравнению с не кодированным распределением битов, показанным на рис. 1.6, а.

Код Грея образуется из суммы по модулю 2 кодовой комбинации натурального кода с ней же, сдвинутой на один разряд вправо, при этом младший разряд сдвинутой комбинации отбрасывается. Пример 4-разрядных кодов – натурального и кода Грея показан в таблице 5.

Таблица 5

Десятичное число	Натуральный двоичный код	Код Грея	
0	0000	0000	
1	0001	0001	
2	0010	0011	
3	0011	0010	
4	0100	0110	Ось симметрии
5	0101	0111	
6	0110	0101	
7	0111	0100	Главная ось симметрии
8	1000	1100	
9	1001	1101	
10	1010	1111	
11	1011	1110	Ось симметрии
12	1100	1010	
13	1101	1011	
14	1110	1001	
15	1111	1000	

Свойства кода Грея:

1. Соседние комбинации отличаются в одной позиции.
2. Смена элементов в каждом разряде происходит в 2 раза реже.
3. В сумме соседних комбинаций по mod 2 число единиц = $m-3$, где m – значность кода.
4. Имеются оси симметрии (отражения) – код является рефлексным.

В заключение отметим, что натуральный двоичный код удобен для различных преобразований, ввода в компьютер, при декодировании и т. п., а код Грея – для передачи по линиям связи.

1.3 Модели каналов

В начале главы, на рис. 1.1, показана структурная схема ЦСПИ, составной частью которого является канал передачи. Чтобы установить терминологию и обозначения, используемые в дальнейшем, рассмотрим более общую модель системы цифровой связи.

1.3.1 Обобщенная схема ЦСПИ

Специфика радиотехнических систем передачи информации (РТСПИ) связана с особенностями распространения радиоволн, которая учитывается при выборе модели канала связи. В остальном процессы, протекающие в РТСПИ, не отличаются от процессов в других системах передачи – системах проводной связи, гидроакустических и оптических. Обратимся к схеме РТСПИ с одним источником и одним получателем (рис. 1.7).

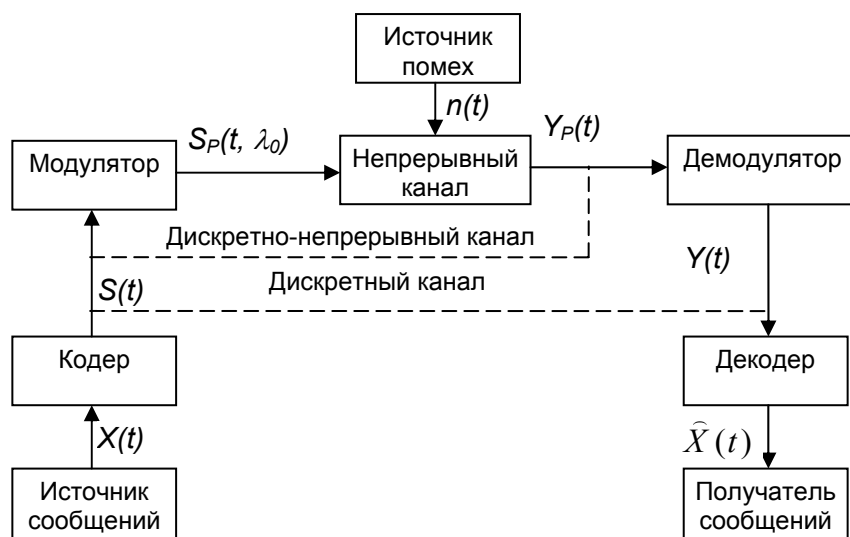


Рис. 1.7. Обобщенная схема цифровой системы передачи.

На этом рисунке использованы обозначения:

$X(t)$ – выход источника сообщений, вход кодера;

$S(t)$ – выход кодера, вход модулятора;

$S_P(t, \lambda_0)$ – выход модулятора (радиосигнал). λ_0 – параметры сигнала;

$Y_P(t)$ – выход канала (радиосигнал), вход демодулятора;

$Y(t)$ – выход демодулятора, вход декодера;

$\hat{X}(t)$ – выход декодера, к получателю сообщений.

Назначение и свойства отдельных блоков схемы были рассмотрены ранее (см. раздел 1.1).

Сигнал $S_P(t, \lambda_0)$ на входе канала имеет параметры $\lambda_0 = \{\omega_0, \varphi, A\}$, которые, в общем случае, зависят от времени. Сигнал $Y_P(t)$ на выходе канала содержит искаженный входной сигнал $S_P(t, \lambda)$ и помеху $n(t)$:

$$Y_P(t) = h_c S_P(t, \lambda_0) + n(t) = S_P(t, \lambda) + n(t) \quad (1.19)$$

Вектор параметров λ , кроме λ_0 , содержит дополнительные составляющие: время запаздывания, доплеровский сдвиг частоты, уменьшение амплитуды и т.п.

Канал будем называть *непрерывным*, если на его входе и выходе действуют непрерывные сигналы; в *дискретном* канале, соответственно, дискретные сигналы. Можно также обозначить *дискретно-непрерывные* и *непрерывно-дискретные* каналы.

1.3.2 Модели непрерывных каналов

Непрерывные каналы можно классифицировать по виду помех и характеру преобразования $S_P(t, \lambda_0)$ в полезный принятый сигнал $S_P(t, \lambda)$. Если ограничиться предположением, что в канале действует аддитивный нормальный белый шум $n(t)$, то непрерывные каналы подразделяются по виду преобразования $S_P(t, \lambda_0)$ в $S_P(t, \lambda)$, т.е. по виду искажений сигнала.

В большинстве радиотехнических систем излучаемые сигналы являются узкополосными:

$$S_P(t, \lambda_0) = A(t) \cos[\omega_0 t + \varphi(t)], \quad (1.20)$$

где $A(t)$ и $\varphi(t)$ – функции, отражающие законы амплитудной и угловой модуляции; ω_0 – несущая частота сигнала, значительно превышающая ширину его спектра. Совокупностью параметров λ_0 является множество $\lambda_0 = \{A(t), \varphi(t), \omega_0(t)\}$.

Искажения излученного сигнала принято рассматривать отдельно для однолучевых и многолучевых каналов. В однолучевых каналах отсутствуют замирания, вызванные интерференцией нескольких сигналов, распространяющихся по различным путям. Многолучевые каналы будут рассмотрены отдельно.

Принятый полезный сигнал по отношению к излученному характеризуется дополнительными параметрами: случайным ослаблением $\alpha(t)$, средним временем запаздывания τ_3 , доплеровским смещением частоты Ω , случайной начальной фазой θ и имеет вид

$$S_P(t, \lambda) = \alpha(t)A(t - \tau_3)\cos[(\omega_0 + \Omega)t + \varphi(t - \tau_3) - \theta]. \quad (1.21)$$

Т.о., совокупность параметров принятого полезного сигнала

$$\lambda = \{A(t), \varphi(t), \omega_0(t), \alpha(t), \tau_3, \Omega, \Theta\}.$$

В зависимости от того, какие параметры принятого сигнала на приемной стороне известны, можно выделить несколько моделей непрерывных каналов:

1. **Гауссовский канал** – канал, в котором помеха имеет вид аддитивного нормального белого шума, а искажения полезного сигнала несут существенны, так как могут быть скомпенсированы. Компенсация искажений возможна, если на приемной стороне дополнительные параметры полностью известны или могут быть измерены достаточно точно (т.е. известна или доступна для измерения импульсная характеристика канала h_c). Поэтому можно считать, что $S_P(t, \lambda_0) = S_P(t, \lambda)$, а выходной сигнал гауссовского канала есть $Y_P(t) = S_P(t, \lambda_0) + n(t)$ (1.22)

2. **Гауссовский канал с неизвестной фазой** сигнала определяется параметрами $\Omega, \tau_3, \alpha(t) = \alpha$, которые постоянны и известны. Начальная фаза неизвестна и обычно считается равномерно распределенной в интервале $(0, 2\pi)$. Такой сигнал хорошо описывает процессы в линии связи на расстоянии прямой видимости.

3. **Канал с амплитудными замираниями** является дальнейшим усложнением канала с неизвестной фазой в предположении, что $\alpha(t)$ – случайная функция времени. Его выходной полезный сигнал имеет вид

$$S_P(t, \lambda) = \alpha(t)A(t - \tau_3)\cos[(\omega_0 + \Omega)t + \varphi(t) - \theta] \quad (1.22)$$

Случайная функция $\alpha(t)$ перемножается с сигналом и поэтому называется *мультипликативной* помехой. Мультипликативную помеху можно рассматривать как функцию, модулирующую амплитуду полезного сигнала. Модуляция приводит к расширению спектра принятого сигнала относительно спектра излученного сигнала. Вследствие этого модель канала (1.22) называют также *каналом с рассеянием энергии* по частоте.

Для задания канала с замираниями требуется описать мультипликативную помеху $\alpha(t)$. Считается достаточным, если указывается одномерная плотность вероятности $W(\alpha)$ и время корреляции τ_k . Если $W(\alpha)$ описывается релеевским законом распределения, то канал называется релеевским или каналом с релеевскими замираниями.

По времени корреляции мультипликативные помехи разделяются на *медленные* и *быстрые*. О медленных замираниях говорят в случае, если время корреляции процесса $\alpha(t)$ значительно превышает интервал наблюдения сигнала. Соответственно, быстрые замирания характеризуются временем корреляции меньшим, чем интервал наблюдения сигнала.

Причинами медленных замираний являются изменения свойств среды распространения радиоволн в зависимости от метеоусловий, времени суток, года, от солнечной активности и т.п. Причиной быстрых замираний может быть, например, наличие в канале нескольких путей распространения радиоволн.

При описании непрерывных каналов важны также ограничения, накладываемые на среднюю или пиковую мощность излучаемых сигналов, на полосу используемых частот.

Т. о., непрерывный канал считается заданным, если указаны мощность сигналов, полоса частот, дано описание моделей помех и искажений сигналов.

1.3.3 Модели дискретных каналов

Дискретными называются каналы, входные и выходные сигналы которых принимают конечное число мгновенных значений. На рис. 1.7 дискретный канал (ДК) включен между кодером и декодером. Очевидно, что дискретный канал всегда содержит непрерывный, так что искажения сигналов и помехи, действующие в дискретном канале, определяются непрерывным каналом.

Переход от дискретных сигналов к непрерывным на передающей стороне осуществляет модулятор; на приемной стороне дискретные сигналы появляются на выходе демодулятора. Поэтому дискретный канал всегда содержит модем.

Таким образом, свойства дискретного канала определяются непрерывным каналом и структурой модема.

ДК задается множеством входных $\{s_i\}$, $i = [1, L_S]$ и выходных $\{y_j\}$, $j = [1, L_y]$ символов длительностью T и условными вероятностями $P(y_j/s_i)$ преобразования входных символов в выходные. Здесь L_S и L_y – объемы алфавитов входных и выходных символов. В общем случае $L_y \geq L_S$, однако, обычно $L_y = L_S = m$.

Для дискретных каналов широко используется представление принятой последовательности символов $Y = \{y_1, y_2, y_3, \dots, y_m\}$ в виде суммы переданной последовательности $S = \{s_1, s_2, s_3, \dots, s_m\}$ и комбинации помехи (вектора ошибки) $E = \{e_1, e_2, e_3, \dots, e_m\}$

$$Y = S \oplus E, \quad (1.23)$$

где знак $\oplus$ означает поразрядное сложение по модулю L_S .

В бинарном случае ($L_S = 2$) нулевой символ вектора ошибки $e_i = 0$ означает, что i -й символ принят правильно ($y_i = s_i$), а $e_i = 1$ – ошибку в приеме ($y_i \neq s_i$).

Классификацию дискретных каналов удобно вести по вектору ошибки E . Наиболее распространены следующие модели.

Канал без памяти – канал, в котором символы e_i являются независимыми случайными величинами. Прием каждого символа в таком канале не зависит от предыстории.

Канал с памятью – канал, в котором прием символа зависит от результата приема предыдущего символа.

Стационарный канал – канал, в котором вероятность ошибочного приема символов не изменяется с течением времени.

Двоичные (бинарные) каналы – каналы, сигналы в которых представляются двоичным кодом с символами 0 и 1. Они задаются с помощью графа, представленного на рис. 1.8.

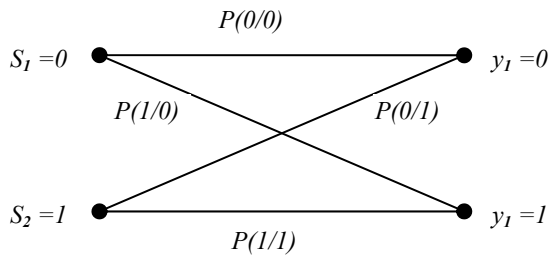


Рис. 1.8.

Вероятности $P(0/0)$ и $P(1/1)$ определяют правильный прием символов 0 и 1 соответственно, а $P(1/0)$ и $P(0/1)$ – вероятности ошибок при приеме символов 0 и 1.

Симметричным двоичным называется канал, в котором вероятности ошибок при приеме 0 и 1 одинаковы, $P(1/0) = P(0/1) = p$. Следовательно, $P(0/0) = P(1/1) = 1 - p = q$, где q – вероятность правильного приема.

Например, вероятность ошибочного приема символа в трехбитовом сообщении $P(000/001) = qq(1-q) = q^2p$.

Учитывая (1.23), можно записать $P(\{y_i\}/\{s_i\}) = P(\{e_i\}/\{s_i\}) = P(\{e_i\})$, $e_i \in \{0, 1\}$, т.е. вероятность трансформации символов определяется вектором ошибок.

Найдем $P_n(t)$ – вероятность t ошибок при приеме n символов. Обозначим через $P_n^*(t)$ вероятность одного сочетания t ошибок в n символах: $P_n^*(t) = p^t(1-p)^{n-t}$. Тогда искомая вероятность определится с учетом всех возможных сочетаний ошибок:

$$P_n(t) = C_n^t P_n^*(t) = n! p^t (1-p)^{n-t} / t!(n-t)! \quad (1.24)$$

Рассмотрим подробнее канал с памятью. Условная вероятность приема сигнала y_i в канале зависит от передачи предыдущих символов: $P(y_i | s_i, s_{i-1}, s_{i-2}, \dots)$. Если обозначить через $(s_{i-1}, s_{i-2}, s_{i-3}, \dots, s_{i-N}) = C_{i-1}$ – состояние канала в $(i-1)$ -й момент времени, то памятью канала N бу-

дем называть максимальное количество предыдущих символов, влияющих на прием текущего символа. Увеличение N не приводит к изменению условной вероятности:

$$P(y_i/s_i, s_{i-1}, \dots, s_{i-N}) = P(y_i/s_i, s_{i-1}, \dots, s_{i-N-j}), \quad j \geq 1 \quad (1.25)$$

Канал задается совокупностью переходных вероятностей $P(C_i / C_{i-1})$ и вероятностью ошибок $P_e^{(i)}$ в i -ом состоянии: $P(y_i, C_i / s_i, C_{i-1})$, $C_i \in L$ – множество состояний канала. Разумно предположить, что передаваемые сигналы s_i и состояние канала C_i – независимы. Тогда условную вероятность приема текущего символа можно записать, как

$$P(y_i, C_i / s_i, C_{i-1}) = P(y_i / s_i, C_{i-1}) P(C_i / C_{i-1}) \quad (1.26)$$

Если текущее состояние канала зависит только от его предыдущего состояния ($N=1$), то канал называют *марковским*, а последовательность состояний описывается простой цепью Маркова.

Рассмотрим *модель Гильберта* – простейшую модель канала с памятью.

1.3.3.1 Модель Гильберта

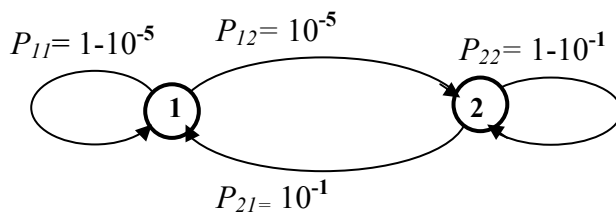
В этой модели канал характеризуется двумя состояниями.

Состояние 1 – хорошее, ошибки отсутствуют.

Состояние 2 – плохое, (например, вероятность ошибок $P_e^{(2)}=0,4$).

Канал можно описать матрицей переходных состояний

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}, \text{ или графом (числовые значения приведены для примера):}$$



Когда $P_{11} \neq P_{22}$, возникают пакеты ошибок.

Вероятности, составляющие полную группу событий:

$$\begin{aligned} P_1 + P_2 &= 1 \\ P_{11} + P_{12} &= 1 \\ P_{22} + P_{21} &= 1 \end{aligned} \quad (1.27)$$

Вероятности пребывания канала в состояниях 1 и 2:

$$\begin{aligned} P_1 &= P_1 P_{11} + P_2 P_{21}; \\ P_2 &= P_2 P_{22} + P_1 P_{12}. \end{aligned} \quad (1.28)$$

Выразим в (1.28) P_2 через P_1 :

$$\begin{aligned} (1-P_1)P_{21} + (1-P_{12})P_1 &= P_1, \text{ откуда} \\ P_1 &= P_{21} / (P_{12} + P_{21}). \end{aligned} \quad (1.29)$$

Аналогично, для состояния 2 получим

$$P_2 = P_{12} / (P_{12} + P_{21}). \quad (1.30)$$

Так как ошибки возникают только в плохом состоянии канала, то среднюю вероятность ошибок найдем, как

$$P_e = P_2 P_e^{(2)} = P_e^{(2)} P_{12} / (P_{12} + P_{21}) \quad (1.31)$$

Найдем длину пакета ошибок как среднее число символов, переданных по каналу в его плохом состоянии:

$$\bar{l}_2 = \sum_{i=1}^{\infty} i P_2(i), \text{ где} \quad (1.32)$$

$P_2(i) = P_{22}^{i-1} P_{21}$ - вероятность того, что возникшее плохое состояние канала будет длиться в течение i символов. Тогда (1.32) переписывается в виде:

$$\bar{l}_2 = \sum_{i=1}^{\infty} i P_{22}^{i-1} P_{21} = 1 \cdot P_{21} + 2 P_{22} P_{21} + 3 P_{22}^2 P_{21} + \dots, \quad (1.33)$$

который представляет собой арифметико-геометрическую прогрессию.

В справочнике [6] приведена сумма такой прогрессии:

$$\sum_{k=0}^{\infty} (a + kr) q^k = \frac{a}{1-q} + \frac{rq}{(1-q)^2}. \text{ Тогда искомое выражение в окончательном виде:}$$

$$\bar{l}_2 = \frac{1}{P_{22}} \frac{P_{21} P_{22}}{(1 - P_{22})^2} = \frac{1}{P_{21}} \quad (1.34)$$

Аналогично найдем среднюю длину интервала между ошибками:

$$\bar{l}_1 = \sum_{i=1}^{\infty} i P_{11}^{(i-1)} P_{12} = \frac{1}{P_{12}} \quad (1.35)$$

Для нашего примера $\bar{l}_{21} = 10$, $\bar{l}_{12} = 10^5$.

1.3.3.2 Характерные искажения в канале

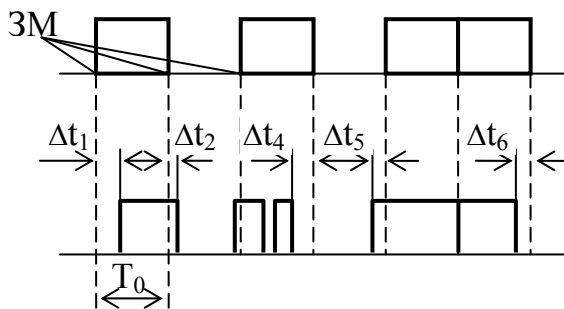


Рис. 1.9. Искажения сигналов

Отметим два характерных вида искажений: *краевые искажения* и *дробления*. На верхней диаграмме рисунка показаны излученные импульсы, а на нижней – сигнал, восстановленный демодулятором. Начало и конец тактового интервала называются значащими моментами (ЗМ). Если фронты восстановленного сигнала не совпадают с ЗМ, то имеют место краевые искажения, которые характеризуются смещениями Δt_i . Если в течение тактового интервала наблюдаются переходы от 0 к 1 и обратно, имеют место дробления сигнала.

1.3.3.3 Регистрация сигналов

Регистрация принятых сигналов осуществляется демодулятором, выход которого является цифровым. В простейшем случае (например, в проводной связи) решение о принятом символе – ноль или единица – производится демодулятором на основе анализа восстановленного сигнала, подверженного краевым искажениям и дроблениям.

Существует два основных вида регистрации сигналов – метод стробирования и метод интегрирования.

1) Метод стробирования

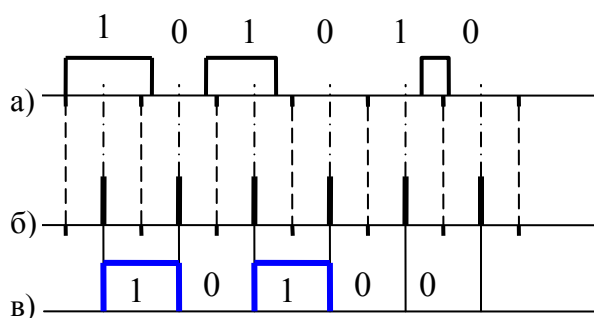


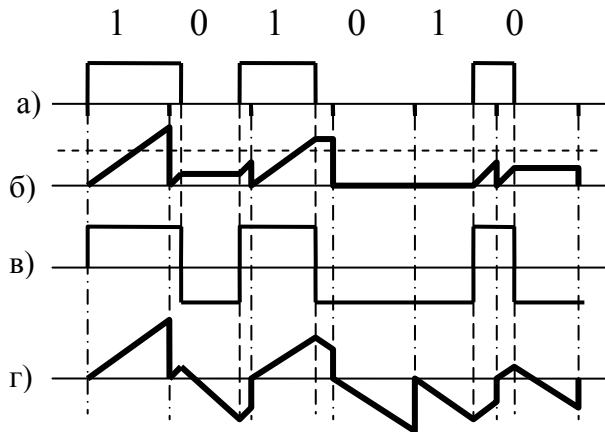
Рис. 1.10. Метод стробирования

возможна как в результате дроблений, так и в результате краевых искажений.

Верхняя диаграмма рис.1.10 а – сигналы на выходе демодулятора. Присутствуют краевые искажения. Для регистрации символов производится взятие отсчетов в середине тактовых интервалов, как показано на второй диаграмме (б). Диаграмма (в) показывает результат регистрации. Очевидно, что метод стробирования будет давать ошибки при наличии дроблений, т.е. всякий раз, когда в центре тактового интервала сигнал искажен помехой. Такая ситуация, впрочем,

2) Метод интегрирования

В этом методе используется накопление (интегрирование) сигналов в течение тактового интервала, т.е. предполагается использование интегратора со сбросом. Сброс интегратора производится в точках ЗМ (рис. 1.11).



При униполярном сигнале (диаграмма а) выход интегратора показан на диаграмме б). Моменты взятия отсчетов выходного напряжения интегратора совпадают с ЗМ. После взятия отсчета интегратор сбрасывается в нуль и процесс накопления возобновляется. Для решения о приеме символа требуется пороговая схема; величина порогового уровня составляет половину максимально возможного напряжения интегратора. Это максимальное напряжение получается при входном сигнале без искажений. Чтобы

Рис. 1.11 Метод интегрирования

избавиться от необходимости выставления порога, обычно используют полярные сигналы (диаграммы в), г)). При этом пороговый уровень равен нулю.

2 Задачи приема и синтез сигналов.

Типичным для теории связи является подход, при котором анализ той или иной системы начинается с приемной стороны. Цель подобной стратегии состоит в разработке оптимального приемного устройства, которое с наилучшим качеством восстановит информацию, содержащуюся в наблюдаемом колебании. Определение оптимального алгоритма обработки, базирующегося на учете специфических свойств переданного сигнала, позволяет в дальнейшем синтезировать оптимальным образом и сам переданный сигнал, т.е. выбрать наилучшим образом методы его кодирования и модуляции. В данной главе будут исследована связь между классическими задачами приема и синтеза ансамблей сигналов. Под классическими задачами здесь понимаются такие, которые базируются на использовании традиционной модели гауссовского канала.

2.1 Общая задача приема, оптимальные решающие правила.

Пусть имеется некоторый источник сообщений, вырабатывающий в каждый конкретный момент времени одно из M возможных сообщений. Каждое из M конкурирующих сообщений передается посредством специфического сигнала, так что имеется множество S из M возможных сигналов: $S = \{s_i(t) : i = 1, 2, \dots, M\}$. На мощность множества S , т.е. число сигналов M , не накладывается никаких ограничений и, если это необходимо, множество S может быть бесконечным. Источник выбирает некоторый определенный сигнал $s_i(t) \in S$ и подает его на вход канала. На приемной стороне (на выходе канала) наблюдается принятое колебание $y(t)$, которое является не точной копией переданного сигнала $s_i(t)$, а результатом трансформации $s_i(t)$, обусловленной искажающим воздействием шумов и помех, присущих любому реальному каналу. Для приемной стороны имеется M конкурирующих гипотез H_i , состоящих в том, какой из M возможных сигналов был в действительности передан и трансформирован каналом в принятое наблюдение $y(t)$, и требуется выбрать только одну из них как истинную. Обозначим результат этого выбора, т.е. решение, через $\hat{H}_j$, означающее, что «решение принято в пользу сигнала с индексом j ». Из данной классической задачи приема вытекает следующая: *какова наилучшая стратегия принятия решения о возможном переданном сообщении (или сигнале), основанная лишь на наблюдении $y(t)$?*

В случае равной вероятности всех сообщений источника (что всегда достигается при правильном проектировании системы) оптимальной стратегией наблюдателя, обеспечивающей минимальную ошибку перепутывания действительно переданного с некоторым другим сигналом, является правило *максимального правдоподобия* (МП). Согласно данному алгоритму после того, как колебание $y(t)$ стало достижимым, решение принимается в пользу того сигнала, для которого вероятность трансформации его каналом в принятое наблюдение $y(t)$ является наибольшим (по сравнению с вероятностями для других сигналов).

В теории связи наиболее распространенной моделью служит канал с аддитивным белым гауссовским шумом (АБГШ–канал), или просто гауссовский канал, в котором переходная вероятность экспоненциально уменьшается с ростом квадрата *Евклидова расстояния* между переданным сигналом и выходным наблюдением:

$$p[y(t)|s(t)] = k \exp\left(-\frac{1}{N_0} d^2(s, y)\right), \quad (2.1)$$

где k – константа, не зависящая от $s(t)$ и $y(t)$; N_0 – спектральная плотность мощности одностороннего белого шума; а Евклидово расстояние между $s(t)$ и $y(t)$ определяется как

$$d(\mathbf{s}, \mathbf{y}) = \sqrt{\int_0^T [y(t) - s(t)]^2 dt} . \quad (2.2)$$

Объяснение чрезвычайной важности гауссовской модели лежит в физической природе многих реальных шумов. Согласно центральной предельной теореме теории вероятности вероятностное распределение суммы большого числа случайных элементарных компонентов, между которыми нет сильной зависимости, и ни один из них не преобладает над другими, аппроксимируется гауссовским законом, особенно при бесконечном числе слагаемых. Но и тепловой шум, и многие другие, типичные для реальных каналов связи, представляют собой результат суммирования значительного числа элементарных токов или напряжений, обусловленных хаотическим движением заряженных частиц (электронов, ионов и т.п.).

Что же касается расстояния между сигналами или колебаниями, то они интерпретируются как вектора, что общепринято во всех информационных и родственных дисциплинах. В случае затруднения понимания связи между сигналами и векторами помощь может оказать следующее простейшее объяснение. Осуществим дискретизацию во времени непрерывного сигнала, т.е. представим сигнал $s(t)$ его отсчетами $s_i = s(iT_s)$, $i = 0, 1, \dots$, взятыми через постоянный интервал T_s . Если вся энергия сигнала сосредоточена в пределах полосы W , а $T_s \leq 1/2W$, то отсчеты s_i полностью определяют исходный непрерывный во времени сигнал $s(t)$. При длительности сигнала T всего имеется $n = T/T_s$ подобных отсчетов и, значит, n -мерный вектор $\mathbf{s} = (s_0, s_1, \dots, s_{n-1})$ полностью описывает сигнал. Прделав подобную же операцию с наблюдением $y(t)$, приходим к его n -мерному векторному эквиваленту $\mathbf{y} = (y_0, y_1, \dots, y_{n-1})$, что позволяет найти Евклидово расстояние между векторами $\mathbf{s}$ и $\mathbf{y}$ согласно теореме Пифагора для n -мерного векторного пространства:

$$d(\mathbf{s}, \mathbf{y}) = \sqrt{\sum_{i=0}^{n-1} [y_i - s_i]^2} .$$

Если устремить T_s к нулю, то вектора $\mathbf{s}$ и $\mathbf{y}$, являвшиеся дискретными эквивалентами сигнала и наблюдения, становятся бесконечной размерности, а значит, восстанавливают $s(t)$, $y(t)$, поскольку фактически отсутствует дискретизация во времени. Одновременно вышеприведенная сумма в правой части равенства (2.2) заменяется (без учета множителя) интегралом, и мы приходим к определению Евклидова расстояния для непрерывных во времени колебаний.

Возвратимся вновь к правилу МП для гауссовского канала. Согласно соотношениям (2.1)–(2.2), правдоподобие сигнала (вероятность того, что он преобразован каналом в наблюдение $y(t)$) уменьшается с увеличением Евклидова расстояния между $s(t)$, $y(t)$. Следовательно, правило МП для гауссовского канала может быть преобразовано в правило *минимума расстояния*:

$$d(\mathbf{s}_j, \mathbf{y}) = \min_i d(\mathbf{s}_i, \mathbf{y}) \Rightarrow \hat{H}_j , \quad (2.3)$$

т.е. решение принимается в пользу сигнала $s_j(t)$, поскольку он наиболее близок (в смысле Евклидова расстояния) к наблюдению $y(t)$ среди всех M конкурирующих сигналов (см. рис.2.1). Другим, более наглядным представлением (2.3), является следующее

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{s} \in S} d(\mathbf{s}, \mathbf{y}) ,$$

где $\hat{\mathbf{s}}$ – оценка принятого сигнала (т.е. сигнал, который считается принятым).

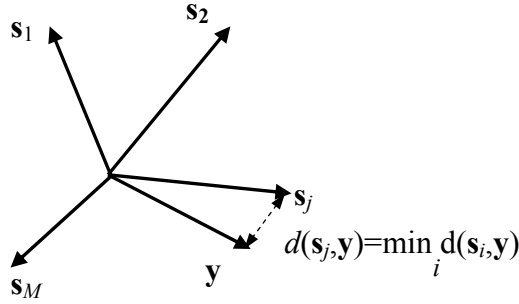


Рис. 2.1. Правило минимума расстояния.

Продолжая оставаться в рамках геометрической интерпретации сигналов, можно ввести длину сигнала $\|\mathbf{s}\|$ как его расстояние относительно начала координат. Тогда из (2.2) следует, что $\|\mathbf{s}\| = d(\mathbf{s}, \mathbf{0}) = \sqrt{E}$, где

$$E = \int_0^T s^2(t) dt \quad (2.4)$$

энергия сигнала. Другой важной геометрической характеристикой является скалярное произведение $(\mathbf{u}, \mathbf{v})$ двух сигналов $u(t)$, $v(t)$:

$$(\mathbf{u}, \mathbf{v}) = \int_0^T u(t)v(t) dt, \quad (2.5)$$

которое снова может трактоваться как предельная форма скалярного произведения двух n -мерных векторов. Эта же характеристика может быть вычислена с помощью длины векторов и косинуса угла α между ними: $(\mathbf{u}, \mathbf{v}) = \|\mathbf{u}\| \|\mathbf{v}\| \cos \alpha$, и, таким образом, скалярное произведение свидетельствует о близости, или *сходстве* сигналов, поскольку, чем ближе сигналы одинаковой длины (энергии) друг к другу, тем меньше $\cos \alpha$ отличается от единицы, и тем больше скалярное произведение. На основании этого скалярное произведение называют также *корреляцией* сигналов.

Рассмотрим несколько иную версию правила минимального расстояния для того, чтобы подчеркнуть особую роль данной характеристики. Раскрыв скобки в (2.2), приходим к соотношению

$$d^2(\mathbf{s}_i, \mathbf{y}) = \int_0^T y^2(t) dt - 2 \int_0^T y(t)s_i(t) dt + \int_0^T s_i^2(t) dt = \|\mathbf{y}\|^2 - 2z_i + \|\mathbf{s}_i\|^2, \quad (2.6)$$

где z_i соответствует корреляции между наблюдением $y(t)$ и i -м сигналом $s_i(t)$

$$z_i = (\mathbf{y}, \mathbf{s}_i) = \int_0^T y(t)s_i(t) dt. \quad (2.7)$$

Первое слагаемое в правой части соотношения (2.6) фиксировано для данного наблюдения и поэтому не влияет на анализируемые расстояния и решение, какой из сигналов был принят. Последний член суммы есть не что иное, как энергия i -го сигнала E_i . Учитывая это, правило минимума расстояния (2.3) может быть сформулировано как следующее правило *максимума корреляции*:

$$z_j - \frac{E_j}{2} = \max_i (z_i - \frac{E_i}{2}) \Rightarrow \hat{H}_j, \quad (2.8)$$

означающее, в частности, что из M возможных сигналов с одинаковой энергией фактически принятым считается тот, который имеет максимум корреляции с наблюдением $y(t)$. Последнее поддается ясной физической трактовке. Предпочтение отдается тому из сигналов, который наиболее подобен наблюдению $y(t)$ в сравнении с остальными при условии, что в качестве критерия сходства принята величина корреляции (скалярное произведение).

Интересно отметить, что даже эти предварительные рассуждения указывают уже более чем явный способ конструирования множества сигналов. Обратимся к рис.2.2, на кото-

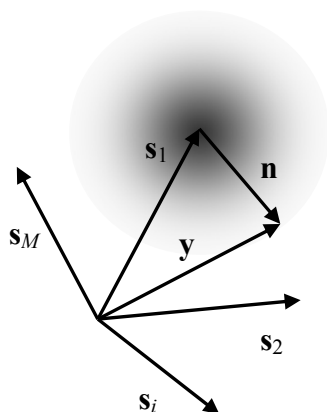


Рис. 2.2. Зашумленное наблюдение и задача выбора сигналов.

ром изображены сигнальные вектора. Предположим, что передавался сигнал s_1 , и что он подвергся искажению в АБГШ канале, следствием чего служит добавление к s_1 вектора шума n . Гауссовский вектор n характеризуется симметричным (сферическим) вероятностным распределением, экспоненциально спадающим с увеличением длины вектора n , что очевидным образом следует из (2.1) после удаления из него сигнала (т.е. при подстановки $s(t) = 0$). Следовательно, вектор наблюдения $y = s_1 + n$ будет случайным образом перемещаться вокруг s_1 , как это показано на рисунке, и тогда, согласно правилу минимума расстояния (2.3), как только y окажется ближе к некоторому другому, чем s_1 , сигналу, то будет принято ошибочное решение. Для минимизации возникновения ошибки подобного рода следует располагать другие сигналы на максимально большом расстоянии от s_1 . Поскольку любой из M сигналов может передаваться равновероятно, т.е. занимать место s_1 , то, очевидно, что все расстояния $d(s_i, s_j), 1 \leq i < j \leq M$, следует делать максимально большим. В случае достаточно большого значения M задача одновременной максимизации всех расстояний не представляется простой, поскольку они могут конфликтовать друг с другом: отодвигая некоторый вектор от другого, возможно приближение первого к некоторому третьему. Учитывая последнее, задача построения множества максимально удаленных друг от друга сигналов (входящая в обширный класс т.н. задач *упаковки*) во многих случаях оказывается достаточно сложной, и пока что не имеет общего решения.

Отметим, что в ранее рассмотренном материале все M сигналов считались полностью *детерминированными*, т.е. полагалось, что на приемной стороне априори известны все параметры сигналов, единственное, что наблюдателю неизвестно, так это какой из конкурирующих M сигналов был принят. Подобная модель в значительной степени адекватна приему сигналов на видеочастоте или когерентному приему на высокой частоте. Однако с некоторыми дополнениями основная идея сохраняет свою значимость и в более сложных ситуациях, таких как некогерентный прием.

Освежив эти основные идеи оптимального приема, теперь можно перейти к рассмотрению специфических задач, уделяя основное внимание аспектам конструирования сигнала

2.2 Передача двоичных данных (детерминированные сигналы).

Строгая зависимость качества приема от расстояния между сигналами наиболее наглядно проявляется в простейшей, но и очень типичной связанной задаче *передачи двоичной информации*, когда по каналу пересылается только одно из $M = 2$ возможных сообщений. Фактически данная ситуация отвечает передаче либо одного бита данных в системе, где отсутствует канальное кодирование, либо одного символа двоичного кода в системе с помехоустойчивым кодированием и жестким решением на приеме и т.п. Обозначая сообщения через нуль и единицу и полагая, что для их передачи используются сигналы $s_0(t)$ и $s_1(t)$ (снова детерминированные), правило решения (2.3) по минимуму расстояния может быть представлено в виде

$$d(s_0, y) \begin{matrix} < \\ > \end{matrix} d(s_1, y), \quad \begin{matrix} \hat{H}_0 \\ \hat{H}_1 \end{matrix} \quad (2.9)$$

где расположение решения о символе прямо указывает, когда принимается то или иное решение. Аналогичным образом может быть переписано и решающее правило (2.8), базирующееся на учете корреляции

$$z = z_0 - z_1 \begin{matrix} > \\ < \end{matrix} \frac{E_0 - E_1}{2}, \quad \begin{matrix} \hat{H}_0 \\ \hat{H}_1 \end{matrix} \quad (2.10)$$

где корреляция $z_i, i = 0, 1$ каждого из сигналов и наблюдения $y(t)$ определяется соотношением (2.7), а энергия сигнала $E_i = \|s_i\|^2, i = 0, 1$ устанавливается (2.4). Оптимальные правила различения (2.9)–(2.10) двух сигналов могут быть наглядно объяснены с помощью их геометрической интерпретации. Два сигнальных вектора s_0 и s_1 всегда лежат в сигнальной плоскости SP . Вектор наблюдения y необязательно попадет на эту плоскость, но близость его к одному или другому сигналу определяется близостью к ним проекции y' вектора y на SP (см. рис. 2.3, а). Следовательно, плоскость SP может быть поделена на две полуплоскости прямолинейной границей, проходящей через начало координат и перпендикулярной к прямой линии, соединяющей сигнальные вектора. Тогда решения $\hat{H}_0$ и $\hat{H}_1$ фактически определяются как попадание проекции y' на соответствующую полуплоскость (см. рис. 2.3, б). Из рис. 2.3, б также следует, что вероятность перепутывания сигналов (вероятность ошибки)

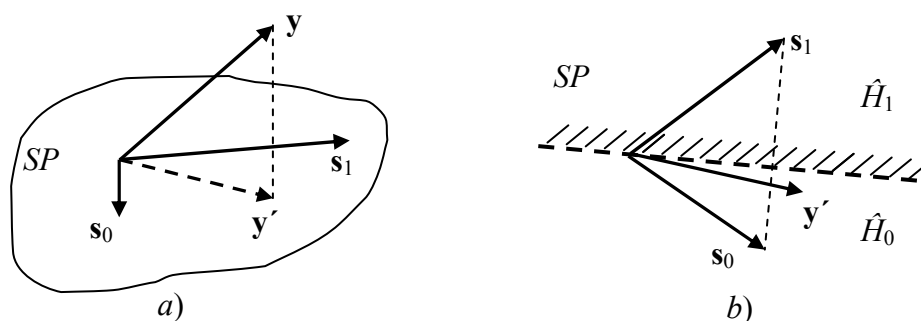


Рис. 2.3. Сигнальная плоскость и полуплоскости решения.

зависит от расстояния между векторами $\mathbf{s}_0$ и $\mathbf{s}_1$ в сравнении с диапазоном случайных флуктуаций $\mathbf{y}'$, обусловленных канальным шумом. Согласно (2.10), фактически принятый сигнал $s_0(t)$ будет ошибочно засчитан за $s_1(t)$, если и только если разность корреляций будет меньше порога $(E_0 - E_1)/2$. Следовательно, вероятность p_{01} подобной ошибки может быть найдена как

$$p_{01} = P\left(z < \frac{E_0 - E_1}{2} \middle| s_0(t)\right) = \int_{-\infty}^{\frac{E_0 - E_1}{2}} W(z|s_0(t)) dz, \quad (2.11)$$

где $P(A|B)$ – условная вероятность события A при условии, что событие B произошло, а $W(z|s_0(t))$ – условная плотность вероятности разности корреляций z из (2.10) при условии, что сигнал $s_0(t)$ действительно был принят. Одним из замечательных свойств гауссовского процесса является то, что любое его линейное преобразование вновь дает гауссовский процесс. Таким образом, поскольку z , согласно (2.7), (2.10), получено как линейное преобразование гауссовского наблюдения $y(t)$, то z также подчиняется гауссовскому распределению

$$W(z|s_0(t)) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(z - \bar{z})^2}{2\sigma^2}\right],$$

интегрирование которого, согласно (2.11), приводит к выражению

$$p_{01} = Q\left(\frac{2\bar{z} - E_0 + E_1}{2\sigma}\right), \quad (2.12)$$

где

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp\left(-\frac{t^2}{2}\right) dt$$

– дополнительная функция ошибок).

Из (2.7), (2.10) могут быть найдены математическое ожидание $\bar{z}$, обусловленное принятым сигналом (черта сверху используется для обозначения усреднения), и дисперсия $\sigma^2 = D\{z\}$. В случае, когда действительно был принят сигнал $s_0(t)$, т.е. $\overline{y(t)} = s_0(t)$, математическое ожидание z будет

$$\bar{z} = \int_0^T \overline{y(t)} [s_0(t) - s_1(t)] dt = E_0 - \rho_{01} \sqrt{E_0 E_1}, \quad (2.13)$$

где

$$\rho_{ij} = \frac{(\mathbf{s}_i, \mathbf{s}_j)}{\|\mathbf{s}_i\| \|\mathbf{s}_j\|} = \frac{1}{\sqrt{E_i E_j}} \int_0^T s_i(t) s_j(t) dt \quad (2.14)$$

– коэффициент корреляции сигналов $s_i(t)$, $s_j(t)$, а E_i , E_j – энергии сигналов. Из (2.14) следует, что в геометрической интерпретации ρ_{01} есть просто косинус угла между сигналами $s_0(t)$, $s_1(t)$ (или сигнальными векторами $\mathbf{s}_0$, $\mathbf{s}_1$), и, значит, характеризует близость, или *подобие* (*сходство*) сигналов.

При нахождении дисперсии σ^2 учтем тот факт, что на нее не оказывает влияние детерминированный компонент наблюдения $y(t)$ (в рассматриваемом случае сигнал $s_0(t)$),

поскольку шум в канале аддитивен. Фактически, тогда можно удалить сигнал из выражения для $y(t)$, полагая $y(t) = n(t)$, где $n(t)$ – белый шум с двусторонней спектральной плотностью мощности $N_0/2$. После приведенных рассуждений нахождение дисперсии корреляции (2.7) наблюдения $y(t)$ и произвольного сигнала $s(t)$ сводится к вычислению

$$\sigma^2 = D\{z\} = \overline{\left\{ \int_0^T n(t)s(t)dt \right\}^2} = \int_0^T \int_0^T \overline{n(t)n(t')} s(t)s(t') dt dt',$$

где квадрат интеграла заменен двойным интегралом с отдельными переменными, изменен порядок интегрирования и усреднения (математическое ожидание суммы равно сумме математических ожиданий) и, окончательно, усреднение применено только к случайному сомножителю под знаком интеграла.

Напомним теперь, что вследствие равномерности спектра белого шума во всем частотном диапазоне его *автокорреляционная функция* (статистическое среднее произведения двух различных временных отсчетов) есть дельта-функция Дирака:

$$\overline{n(t)n(t')} = (N_0/2)\delta(t-t').$$

Другими словами, любые два отсчета белого шума, не совпадающие во времени, являются некоррелированными. Используя последний результат в выражении для дисперсии, а также *фильтрующее* свойство дельта-функции

$$\int_0^T s(t')\delta(t'-t)dt' = s(t),$$

приходим к выражению

$$\sigma^2 = \frac{N_0}{2} \int_0^T s^2(t)dt = \frac{N_0 E}{2}, \quad (2.15)$$

где E – энергия сигнала $s(t)$.

Как следует из (2.10) и (2.7), в рассматриваемой ситуации в (2.15) следует подставить $s(t) = s_0(t) - s_1(t)$, т.е. E представляет собой энергию E_d разностного сигнала $s_0(t) - s_1(t)$. Учет последнего дает

$$E_d = \int_0^T [s_0(t) - s_1(t)]^2 dt = d^2(\mathbf{s}_0, \mathbf{s}_1) = E_0 + E_1 - 2\rho_{01}\sqrt{E_0 E_1}. \quad (2.16)$$

Принимая во внимание геометрическую интерпретацию коэффициента корреляции и энергии, последнее соотношение представляет собой теорему косинусов из школьной математики.

Окончательно, подстановка (2.13), (2.15) и (2.16) в (2.12) дает

$$p_{01} = Q\left(\sqrt{\frac{d^2(\mathbf{s}_0, \mathbf{s}_1)}{2N_0}}\right). \quad (2.17)$$

Поскольку задача нахождения вероятности ошибочного приема симметрична, то аналогичный результат будет получен и для вероятности p_{10} перепутывания $s_1(t)$ с $s_0(t)$. Учитывая последнее, полная (безусловная) вероятность ошибочного приема P_e не зависит априори от вероятности w поступления в канал сигнала $s_0(t)$ и устанавливается соотношением

$$P_e = wp_{01} + (1-w)p_{10} = Q\left(\sqrt{\frac{d^2(\mathbf{s}_0, \mathbf{s}_1)}{2N_0}}\right). \quad (2.18)$$

Из (2.18) очевидным образом вытекает, что единственным путем достижения высокой

достоверности передачи данных является увеличение расстояния $d(\mathbf{s}_0, \mathbf{s}_1)$ между сигналами до максимально возможного значения. Ясно, что увеличение $d(\mathbf{s}_0, \mathbf{s}_1)$ может быть достигнуто за счет большей энергии сигналов, или длины соответствующих векторов, как это демонстрирует соотношение (2.16). Однако какова будет *оптимальная* пара сигналов, если ресурс прямого решения лимитирован, т.е. энергия сигналов ограничена заранее? Рассмотрим первоначально типичный случай сигналов равных энергий $E_0 = E_1 = E$, когда интенсивность сигналов не используется как индикатор передаваемого сообщения. Тогда правило принятия решения (2.10) сводится к сравнению значений корреляций z_0 и z_1 или, что эквивалентно, определению знака их разности

$$z = z_0 - z_1 \stackrel{\hat{H}_0}{>} \stackrel{\hat{H}_1}{<} 0.$$

Очевидно, что для максимизации расстояния между двумя векторами фиксированной длины следует выбирать их *противоположными*, как это показано на рис. 2.4, а. Тогда угол между векторами $\mathbf{s}_0$ и $\mathbf{s}_1$ составит величину $\alpha = \pi$, а $\cos \alpha = \rho_{01} = -1$ и, значит, $d(\mathbf{s}_0, \mathbf{s}_1) = 2\sqrt{E}$, что преобразует соотношение (2.18) в следующее

$$P_{e,a} = Q\left(\sqrt{\frac{2E}{N_0}}\right), \quad (2.19)$$

обеспечивающее минимально достижимую вероятность ошибки при передаче двоичных данных сигналами с фиксированной энергией E . Такие устройства, как *коррелятор* и *согласованный фильтр*, которые часто упоминаются в дальнейшем, используются для непосредственного вычисления корреляции z , а параметр $q = \sqrt{2E/N_0}$ есть ничто иное, как *отношение сигнал-шум* на выходе упомянутых устройств.

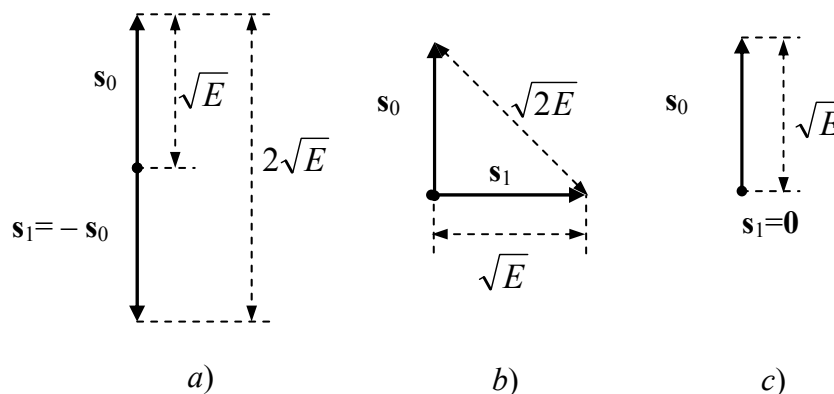


Рис. 2.4. Варианты выбора пар двоичных сигналов.

Таким образом, оптимальной сигнальной парой служат противоположные сигналы вида $s_1(t) = -s_0(t)$. *Бинарная фазовая манипуляция* БФМ (*binary phase shift keying* (BPSK)) представляет собой их практическую реализацию, широко используемую в цифровых системах передачи данных. При этом нулевой символ данных передается радиосигналом с начальной фазой, равно нулю, а тот же самый сигнал, но с начальной фазой высокочастотного заполнения, равной π , используется для передачи единицы.

Для определения критичности в выборе сигнальной пары сравним БФМ с другим популярным методом передачи двоичных данных. Хотя БФМ является наилучшим из возмож-

ных способов передачи двоичной информации, его применение основано на фазовом различии двух сигналов, носителей информации, и, следовательно, требует точного знания на приемной стороне текущего значения фазы несущей частоты. Практическая реализация этого достигается путем применения специального устройства восстановления несущей, что иногда рассматривается, как нежелательное усложнение приемной схемы. Избежать подобного усложнения удастся при использовании другого способа манипуляции – *частотной ЧМ (frequency shift keying (FSK))*, при которой сообщения 0 и 1 передаются сигналами на различных частотах. Типичным является такой выбор несущих частот, при котором сигналы оказываются *ортогональными*, т.е. $\cos \alpha = \rho_{01} = 0$, $d(s_0, s_1) = \sqrt{2E}$ (см. рис. 2.4, *b*). Подстановка этих значений в (2.18) дает

$$P_{e,o} = Q\left(\sqrt{\frac{E}{N_0}}\right). \quad (2.20)$$

Сравнение результатов (2.20) и (2.19) демонстрирует, что для ортогональной пары (ЧМ) обеспечение вероятности ошибочного приема, аналогичной случаю использования противоположных сигналов (БФМ), достигается только при двукратном увеличении энергии сигналов. Иными словами, ортогональные сигналы проигрывают 3 дБ противоположным в уровне необходимой энергии.

Существует еще один, достаточно старый, способ передачи двоичной информации, до сих пор использующийся на практике: *амплитудная модуляция АМ (amplitude shift keying (ASK))*, при которой символ данных «1» передается сигналом $s_1(t) = s(t)$ с энергией $E_1 = E$, а символ «0» – паузой, т.е. $s_0(t) = 0$, $E_0 = 0$. В этом случае (см. рис. 2.5, *c*) $d(s_0, s_1) = \sqrt{E}$, а вероятность ошибки (2.18) становится равной

$$P_{e,as} = Q\left(\sqrt{\frac{E}{2N_0}}\right). \quad (2.21)$$

Сравнивая последний результат с (2.19), можно сделать вывод о том, что АМ требует в 4 раза (6 дБ) большей энергии, чем в случае БФМ, для обеспечения одинаковой достоверности приема. Последнее утверждение справедливо тогда, когда накладывается ограничение на пиковую энергию. На практике, как правило, ограничение накладывается на среднюю энергию. А поскольку при АМ энергия не излучается при передаче нулевых символов, то при равновероятных сообщениях «0» и «1» средняя энергия определится, как $(E_0 + E_1)/2 = E/2$. Тогда при условии одинаковой вероятности ошибки АМ проигрывает БФМ в средней энергии только в два раза. Значит, для АМ по отношению к БФМ характерны такие же энергетические потери, как и в случае применения ЧМ, т.е. 3 дБ.

2.2.1 Прием сигналов БФМ

Запишем сигнал БФМ в виде

$$s(t) = \begin{cases} s_1(t) = S_0 \sin(\omega_0 t + \varphi_1), & \text{если передается 1} \\ s_0(t) = S_0 \sin(\omega_0 t + \varphi_2), & \text{если передается 0} \end{cases}$$

$$|\varphi_1 - \varphi_2| = \pi$$

Замечание о записи амплитуды сигнала.

В самом общем виде сигнал описывается выражением $s(t) = A \cos \omega t$, где A – амплитуда, т.е. максимальное значение сигнала. Среднеквадратичное значение амплитуды $A_{ср}$ в $\sqrt{2}$ меньше амплитуды A , т.е. можно записать $s(t) = A_{ср} \sqrt{2} \cos \omega t = \sqrt{2} A_{ср} \cos \omega t$.

$A_{ср}^2$ представляет собой среднюю мощность P , рассеиваемую на сопротивлении 1 Ом, сле-

довательно

$$s(t) = \sqrt{2P} \cos \omega t.$$

Заменяя P на E/T , получаем:

$$s(t) = \sqrt{\frac{2E}{T}} \cos \omega t \quad (2.22)$$

Конец замечания.

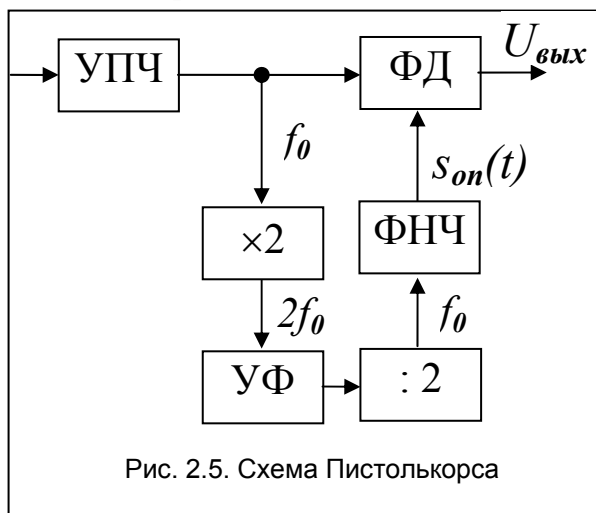
Очевидно, различение таких сигналов должно производиться фазовым детектором, для которого требуется формирование опорного напряжения. Если полярность опорного напряжения изменится на противоположную, полярность выходного сигнала фазового детектора также изменится на обратную. Это явление называется *обратной работой* фазового детектора, и оно становится причиной ошибок демодуляции. Основная трудность реализации демодулятора сигналов БФМ состоит в создании синхронного с ними опорного колебания, имеющего неизменную начальную фазу.

Возможны два пути решения этой проблемы: использование высокостабильного генератора в качестве памяти фазы высокочастотного колебания и формирование опорного колебания с помощью обработки входного сигнала.

Элементарные расчеты показывают, что для обеспечения длительного (более 20 с) сеанса связи с хранением в приемнике начальной фазы требуется генератор с чрезвычайно высокой относительной стабильностью (лучше 10^{-10}). Кроме того (и это может оказаться главной причиной), из-за нестабильности среды распространения начальная фаза принимаемого колебания испытывает флуктуации, которые невозможно учесть в данном методе.

Второй путь – создание опорного колебания из входного сигнала – также содержит проблемы, из которых основной является отсутствие несущей в спектре фазоманипулированного колебания при равновероятных символах. Тем не менее, было предпринято множество попыток создания схем формирования опорного колебания из входного, которые кратко рассматриваются далее.

Одной из первых в 1933 году советским ученым А. А. Пистолькорсом была предложена схема (рис. 2.5), носящая его имя.



Работа схемы происходит на промежуточной частоте f_0 , что, впрочем, совершенно непринципиально. С помощью умножителя частоты на два ($\times 2$) колебание на его выходе не зависит от манипуляции фазы входного сигнала. Узкополосный резонансный фильтр (УФ), настроенный на частоту $2f_0$, позволяет получить на своем выходе почти синусоидальное колебание, в значительной степени свободное от помех. В качестве узкополосного фильтра более удобно использовать схему ФАПЧ, которая способна отслеживать флуктуации частоты, вызванные нестабильностью передатчика и влиянием канала. После деления частоты на два ($:2$) получается необходимое опорное колебание с частотой f_0 без фазовой манипуляции и с малым уровнем помех. Фазовращатель ФВ служит для компенсации аппаратурных фазовых набегов.

Однако, исследования показали, что в зависимости от начальных условий, фаза сформированного опорного колебания может принимать два значения относительно фазы входного сигнала – “0” и “ π ”, что приводит к обратной работе ФД. Кроме того, интенсивные помехи также могут вызывать перескок фазы опорного напряжения на π . Наличие делителя на два увеличивает вероятность перескоков фазы.

Приведенная на рис. 2.6 схема предложена в 1937 году советским ученым В.И. Сифоровым. В этой схеме фазовая манипуляция также снимается с помощью удвоителя частоты, но в качестве узкополосного фильтра используется следящая схема (ФАПЧ), состоящая из дискриминатора фазы (ДФ), фильтра нижних частот (ФНЧ), управляемого генератора (УГ) и второго удвоителя частоты ($\times 2$). В этой схеме отсутствует удвоитель частоты, так как управляемый генератор УГ работает на частоте f_0 , а сравнение фаз выполняется в фазовом дискриминаторе ДФ на удвоенной частоте. Однако отсутствие делителя не устраняет неоднозначности фазы полученного опорного колебания.

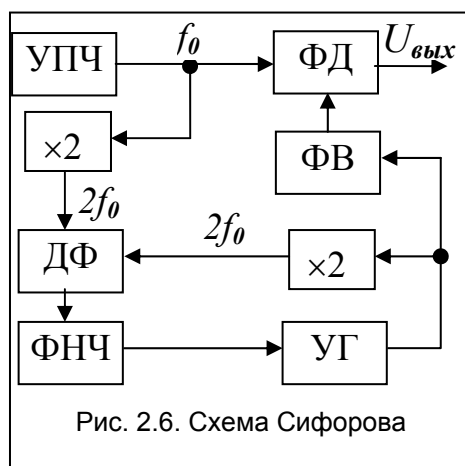


Рис. 2.6. Схема Сифорова

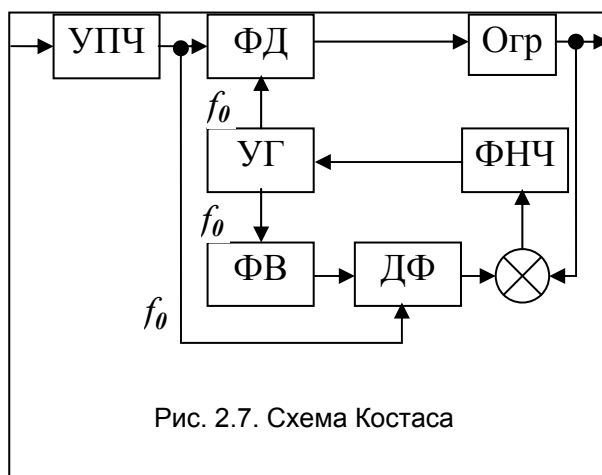


Рис. 2.7. Схема Костаса

П
редло
жен-
ная в
1957
году
амер-
и-
кан-
ским
уче-

ным Д. Костасом схема формирования опорного напряжения показана на рис. 2.7. Все элементы схемы работают на частоте f_0 , а фазовая манипуляция снимается умножением протектированного сигнала на входной в умножителе, выход которого после низкочастотной фильтрации управляет частотой генератора УГ. Эта схема проще в реализации, но, как и предыдущие, обладает склонностью к обратной работе.

Таким образом, прием оптимальных противоположных сигналов наталкивается на принципиально неустранимую трудность – формирование опорного напряжения из принимаемого сигнала, в котором отсутствует составляющая несущей частоты.

Решают эту проблему введением дополнительного канала для передачи несущей частоты – так называемого пилот-сигнала. Если такое решение недопустимо по каким-либо причинам (сложность аппаратуры, стоимость и т.п.), выход состоит в использовании относительной фазовой манипуляции – ОФМ.

2.2.2 Прием сигналов ОФМ

Относительная бинарная фазовая манипуляция – ОБФМ (англ.: differential phase-shift keying – DPSK) – состоит в относительной кодировке передаваемых символов и относительного декодирования, которое может быть когерентным и некогерентным.

Ценность принципа ОФМ состоит в том, что он как бы трансформирует канал связи с переменными параметрами в канал с почти постоянными параметрами, так как на отрезке двух сравниваемых посылок заметных изменений в среде распространения не происходит (при достаточно коротких посылках).

Правило кодирования основано на применении относительных кодов (см. раздел 1.5): текущий передаваемый символ x_k меняет свою фазу относительно предыдущего при информационном символе m_k “1” (М-код) или “0” (S-код).

$$\begin{aligned} x_k &= x_{k-1} \oplus m_k \\ x_k &= x_{k-1} \oplus m_k \end{aligned} \quad (2.23)$$

Декодирование принятых символов производится сравнением фазы текущего символа с фазой предыдущего.

В общем случае М-ичной передачи приемник должен определять положение векторов текущего и предыдущего сигналов и измерять угол между ними (рис. 2.8). Если частота ω_0 передаваемого сигнала известна, приемник вычисляет координаты поступающего сигнала, сравнивая его с локально генерируемыми сигналами $A \cdot \cos \omega_0 t$ и $A \cdot \sin \omega_0 t$. Начальную фазу принимаемого сигнала при этом знать не обязательно.



Рис. 2.8. Сигнальное пространство для М-ичной ОФМ

Если определить набор сигналов ОБФМ следующим образом

$$\begin{aligned} x_1(t) &= A \cos(\omega_0 t + \varphi) & 0 \leq t \leq T, \\ x_2(t) &= A \cos(\omega_0 t + \varphi \pm \pi) & 0 \leq t \leq T, \end{aligned}$$

то решение основывается на разности фаз между принятыми сигналами. Таким образом, при передаче сигналов ОБФМ каждый бит в действительности передается парой двоичных сигналов:

$$\begin{aligned} s_1(t) &= (x_1, x_1) \quad \text{или} \quad (x_2, x_2) & 0 \leq t \leq 2T, \\ s_2(t) &= (x_1, x_2) \quad \text{или} \quad (x_2, x_1) & 0 \leq t \leq 2T \end{aligned} \quad (2.24)$$

Запись (x_i, x_j) обозначает сигнал $x_i(t)$, за которым следует сигнал $x_j(t)$. Первые T секунд каждой пары – это в действительности последние T секунд предыдущей пары. Два бита передается тремя двоичными сигналами, $(m-1)$ битов – m двоичными сигналами. Отметим, что $x_1(t)$ и $x_2(t)$ – противоположные сигналы. Тогда корреляцию между $s_1(t)$ и $s_2(t)$ можно записать, как

$$z(2T) = \int_0^{2T} s_1(t) s_2(t) dt = \int_0^T [x_1(t)]^2 dt - \int_0^T [x_1(t)]^2 dt = 0.$$

Следовательно, каждую пару сигналов ОБФМ можно рассматривать как ортогональный сигнал длительностью $2T$ секунд с энергией $E_b = 2E_p$, где E_p – энергия сигнала $x_i(t)$. Приемник таких сигналов, как ортогональных, изображен на рис. 2.9.

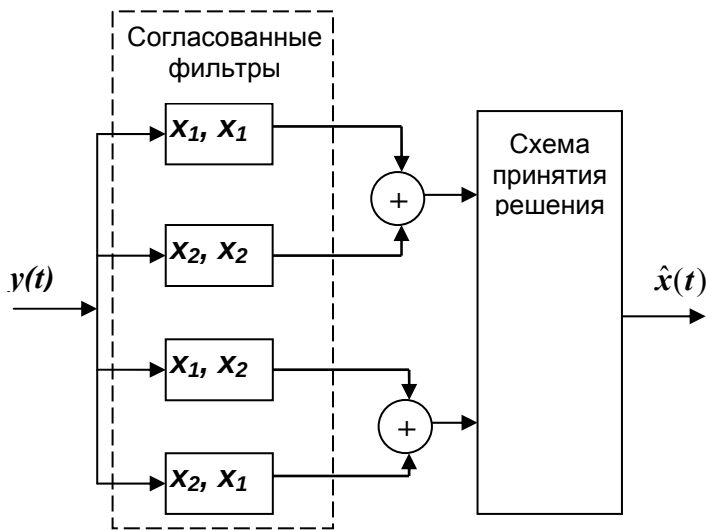


Рис. 2.9. Дифференциально-когерентный приемник ОБФМ

Используя выражение для вероятности ошибки приема ортогональных сигналов (2.20), P_E для сигналов ОБФМ можно записать в виде

$$P_E = 0,5 \exp(-E_b / N_0) \quad (2.25)$$

На рис. 2.10 представлен приемник, построенный по когерентной схеме.

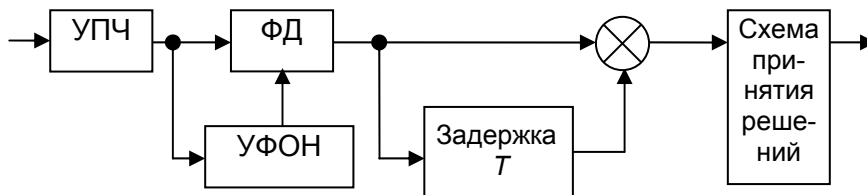


Рис. 2.10. Квазикогерентный приемник ОБФМ.

Но так как устройство формирования опорного напряжения УФОН не полностью освобождает опорный сигнал от шумов, правильнее назвать его квазикогерентным. Тем не менее, каждый из двух импульсов сигнала ОБФМ детектируется когерентно, и к ним можно применить формулу расчета вероятности ошибки, полученную для когерентного приема противоположных сигналов (2.19).

Но ошибка в приеме любого двоичного сигнала влечет за собой ошибку в приеме следующего: если на k -ой позиции сигнал принят верно, а на $(k-1)$ -й неверно и наоборот, то общая вероятность ошибки с учетом неполного устранения шумов в опорных сигналах теперь будет

$$P_E = 2P_{IE} (1 - P_{IE}) + P_{OP},$$

где $P_{IE} = Q(h\sqrt{2})$ – вероятность ошибки приема одного двоичного сигнала, $h^2 = E_b / N_0$ – отношение сигнал/шум, P_{OP} – ошибка за счет шумов опорного канала. Поскольку влияние опорного канала может быть значительно ослаблено за счет сужения полосы фильтрации, им можно пренебречь. Таким образом, полная вероятность ошибки будет

$$P_E = 2P_{IE} (1 - P_{IE}) = 2Q(h\sqrt{2})[1 - Q(h\sqrt{2})] = 2[Q(h\sqrt{2}) - Q^2(h\sqrt{2})]$$

Так как вероятность ошибки $P_E \ll 1$, то $Q(h\sqrt{2}) \rightarrow 0$, и

$$P_E \approx 2Q(h\sqrt{2}), \text{ или,} \quad (2.26)$$

используя разложение гауссовой функции ошибок $Q(h\sqrt{2})$ в ряд, получим

$$D_E \approx \frac{1}{h\sqrt{\pi}} \exp(-h^2). \quad (2.27)$$

Существуют и другие способы приема ОБФМ-сигналов. Автокорреляционный приемник, представленный рис. 2.11 а, можно классифицировать, как когерентный, поскольку опорным сигналом фазового детектора ФД служит предыдущий двоичный символ, а параметры канала за время $2T$ не меняются. Входные сигналы можно записать в виде

$$x_j(t) = m_j(t) + n_j(t);$$

$$x_{j-1}(t-\tau_0) = m_{j-1}(t-\tau_0) + n_{j-1}(t-\tau_0),$$

где $m_j(t)$ – двоичный сигнал, сформированный в передатчике, $n_j(t)$ – гауссовский шум. Наличие шума в обоих импульсах сигнала ОБФМ приводит к появлению шумовых составляющих выходного сигнала, вызванные произведением $x_j(t) \cdot x_{j-1}(t-\tau_0)$. Вероятность ошибки такого приема имеет вид (2.25).

Квадратурный приемник (рис. 2.11 б) определяет синфазность или противофазность текущего двоичного символа относительно предыдущего. Так, если $x_j(t)$ и $x_{j-1}(t-\tau_0)$ синфазны, результаты произведений $z_1(t)$ и $z_2(t)$ будут положительны и наоборот.

Вероятность ошибки этого способа приема, относящегося к некогерентным, также определяется формулой (2.25): $P_E = 0,5 \exp(-h^2)$.

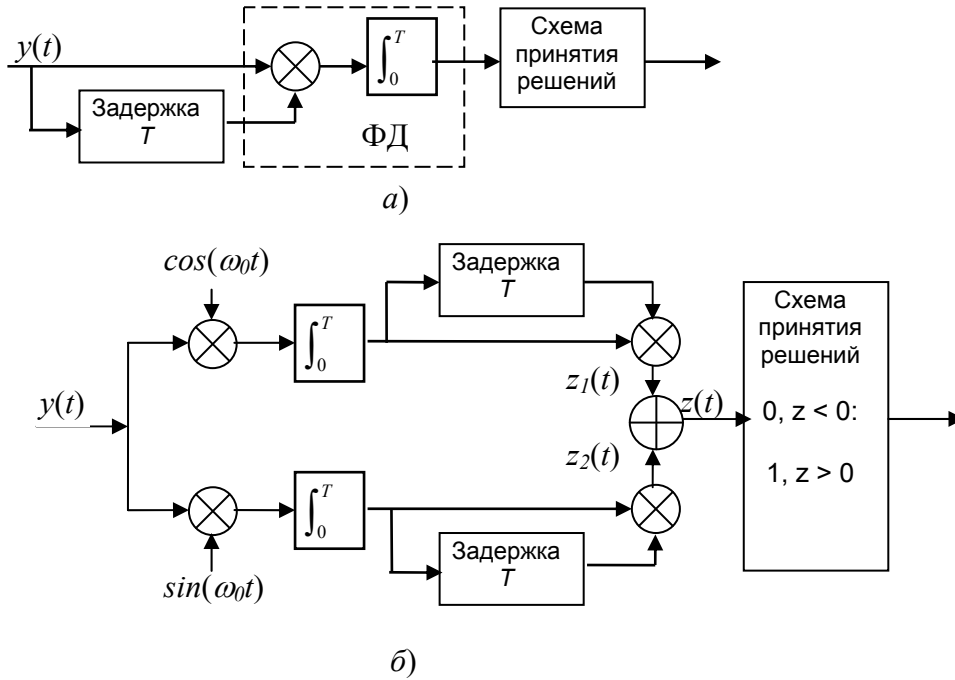


Рис. 2.11. Приемники ОФМ-сигналов: а) автокорреляционный; б) квадратурный

2.3 Передача M-ичных данных

При возможном числе сигналов $M > 2$ вероятность $p_{1,e}$ перепутывания действительно принятого $s_1(t)$ с одним из $M-1$ ошибочных сигналов $s_j(t)$, $j = 2, 3, \dots, M$, определяется, согласно решающим правилам (2.3) и (2.8), как

$$p_{1,e} = P(d^2(\mathbf{s}_1, \mathbf{y}) \neq \min_i d^2(\mathbf{s}_i, \mathbf{y}) | s_1(t)) = 1 - P(z_1 - \frac{E_1}{2} = \max_i (z_i - \frac{E_i}{2}) | s_1(t)).$$

Определение точной оценки этой вероятности состоит в интегрировании условной – при истинности сигнала $s_1(t)$ – совместной плотности вероятности всех M корреляций по

всей области выполнения неравенства $z_1 \geq z_i - \frac{E_i - E_1}{2}$, $i = 1, 2, \dots, M$. Этот M – кратный интеграл в общем случае, т.е. без допущения о некоторых особых свойствах множества сигналов, не может быть упрощен. Однако весьма полезная и точная оценка сверху вероятности $p_{1,e}$ может быть получена с использованием аддитивной границы. Пусть событие A_j состоит в том, что наблюдение $y(t)$ оказывается ближе к ошибочному сигналу $s_j(t)$, $j = 2, 3, \dots, M$, чем к истинному $s_1(t)$. Тогда принятие $s_1(t)$ за какой-то (неважно какой) ошибочный сигнал есть объединение всех A_j .

Согласно аддитивной границе вероятность объединения событий никогда не может быть больше суммы их вероятностей, т.е.

$$p_{1,e} = P(A_2 \cup A_3 \cup \dots \cup A_M) \leq \sum_{j=2}^M P(A_j).$$

С другой стороны, вероятность $P(A_j)$, как следует из определения события A_j , в точности совпадает с вероятностью перепутывания только между двумя сигналами $s_1(t)$ и $s_j(t)$. Значение этой вероятности определяется соотношением (2.12) после соответствующей замены номеров сигналов:

$$P(A_j) = p_{1j} = Q\left(\sqrt{\frac{d^2(\mathbf{s}_1, \mathbf{s}_j)}{2N_0}}\right).$$

Подстановка последнего соотношения в предыдущее неравенство приводит к искомой оценке

$$p_{1,e} \leq \sum_{j=2}^M Q\left(\sqrt{\frac{d^2(\mathbf{s}_1, \mathbf{s}_j)}{2N_0}}\right).$$

Аналогичный результат (с необходимой подстановкой новых числовых индексов) получится и в предположении, что вместо $s_1(t)$ истинным является сигнал $s_i(t)$, так что, с учетом равной априорной вероятности всех M сигналов, аддитивная верхняя граница для полной (безусловной) вероятности ошибки примет вид

$$P_e = \frac{1}{M} \sum_{i=1}^M \sum_{\substack{j=1 \\ j \neq i}}^M p_{i,e} \leq \frac{1}{M} \sum_{i=1}^M \sum_{\substack{j=1 \\ j \neq i}}^M Q\left(\sqrt{\frac{d^2(\mathbf{s}_i, \mathbf{s}_j)}{2N_0}}\right). \quad (2.28)$$

Первым заслуживающим внимание фактом, который непосредственно следует из (2.28), является то, что при $M = 2$ данное неравенство превращается в равенство (2.18). Другое замечание связано с поведением (2.28) в асимптотике при росте отношения сигнал-шум. В самом деле, дополнительная функция ошибки $Q(x)$ уменьшается как $\exp(-x^2/2)$ при достаточно большом x и даже незначительное его увеличение может уменьшить $Q(x)$ до пренебрежимо малого уровня по сравнению с начальным значением. Вследствие этого при достаточно большом отношении сигнал-шум только близко расположенные сигнальные пары могут вносить заметный вклад в значение суммы в (2.28), и если $d_{\min}$ есть минимальное расстояние на множестве всех возможных пар, встречающееся $n_{\min}$ раз, то оценка (2.28) асимптотически трансформируется к виду

$$P_e \approx \frac{n_{\min}}{M} Q \left(\sqrt{\frac{d_{\min}^2}{2N_0}} \right), \quad \frac{d_{\min}^2}{2N_0} \gg 1. \quad (2.29)$$

Приближение (2.29) указывает, прежде всего, на асимптотическую сходимость аддитивной границы к истинной вероятности ошибки при увеличении отношения сигнал-шум. Для физического объяснения данного факта обратимся вновь к рис. 2.3 и отметим, что при малом уровне шума только те сигнальные вектора, которые расположены вблизи от истинного, могут быть ошибочно приняты за последний.

Соотношение (2.29) лежит в основе одной из возможных и наиболее важных постановок задачи синтеза ансамбля сигналов, требующей *максимизации минимума расстояния в множестве M сигналов*. Подобная задача в геометрической интерпретации эквивалентна упаковке M векторов таким образом, чтобы ближайшая их пара находилась на максимально достижимом расстоянии, т.е. $d_{\min} = \max$. При этом на созвездие сигналов (векторов) могут быть наложены различные ограничения. Прежде всего, следует предусмотреть определенные энергетические ограничения, учитывая доступные ресурсы мощности или энергии. Если фиксировать только среднюю энергию сигналов $\bar{E} = \sum_{i=1}^M E_i / M = \text{const}$, то сигнальные вектора могут иметь различную длину и процедуру выбора сигналов можно назвать *объемной упаковкой*. Тем не менее, часто при решении задачи отображения сообщений в сигналы выдвигается требование исключения влияния энергетического параметра, т.е. требование одинаковой $E = E_i = \text{const}$, $i = 1, 2, \dots, M$ энергии всех сигналов. В этом случае все сигнальные вектора имеют одинаковую длину, и значит, лежат на сферической поверхности, откуда и следует название задачи – *сферическая упаковка*.

Другим типичным ограничением при синтезе сигналов является размерность n_s сигнального пространства, внутри которого и осуществляется их упаковка. Физическая сущность этого ограничения снова обусловлена лимитом доступного ресурса, но теперь спектрального. Для объяснения существующей связи рассмотрим первоначально видеосигналы и предположим, что полный (двусторонний) диапазон частот и временной интервал, отведенные для всех M сигналов, ограничены значениями W_t и T_t соответственно. Первое из этих ограничений учитывает экономию полосы, тогда как второе отражает желание передавать данные с приемлемой скоростью $R = \log M / T_t$. Тогда, согласно теореме отсчетов, имеется около $W_t T_t$ независимых отсчетов, доступных для построения M сигналов, причем каждый из сигналов трактуется как вектор в пространстве размерности $n_s = W_t T_t$. Определенная осторожность в оценке числа независимых отсчетов объясняется невозможностью концентрации энергии любого сигнала в пределах конечного и во времени и по частоте интервала. Однако в первом приближении этим фактом можно пренебречь.

Для того, чтобы охватить и случай полосных (радио) сигналов, обратимся к общей модели подобного сигнала

$$s(t) = S(t) \cos[2\pi f_0 t + \gamma(t)], \quad (2.30)$$

где $S(t)$ – вещественная огибающая сигнала, отражающая закон модуляции амплитуды, $\gamma(t)$ описывает закон фазовой модуляции, а f_0 – несущая частота. Используя тригонометрическое преобразование для косинуса суммы углов, (24) можно представить в виде

$$s(t) = S_I(t) \cos 2\pi f_0 t - S_Q(t) \sin 2\pi f_0 t, \quad (2.31)$$

где $S_I(t) = S(t) \cos \gamma(t)$ и $S_Q(t) = S(t) \sin \gamma(t)$ – квадратурные компоненты сигнала. Поскольку $S(t)$ и $\gamma(t)$ являются видеосигналами, то данное утверждение справедливо и в

отношении $S_I(t), S_Q(t)$. Последнее означает, что при переносе на несущую частоту любой радиосигнал исчерпывающе описывается двумя независимыми низкочастотными квадратурными компонентами. Следовательно, при синтезе полосных сигналов могут использоваться в два раза больше независимых координат (отсчетов), чем в случае видеосигналов при одинаковых значениях частотно-временного произведения и, значит, $n_s = 2W_i T_i$.

Теперь в общем виде задача выбора множества сигналов может быть сформулирована следующим образом: найти в пространстве заданной размерности n_s созвездие из M точек или векторов, удовлетворяющее энергетическим ограничениям и обладающее максимально возможным минимумом расстояния между точками $d_{\min} = \max$. Данная задача может быть переформулирована в дуальную: найти в пространстве заданной размерности n_s созвездие из M точек или векторов с гарантированным минимальным расстоянием $d_{\min}$, которое обеспечивает минимизацию энергетических затрат, либо в терминах средней энергии $\bar{E} = \min$ (объемная упаковка), либо равной для всех сигналов $E = \min$ (сферическая упаковка). Простейшая версия этой задачи ($n_s = 1$) отвечает случаю амплитудной (АМ) модуляции

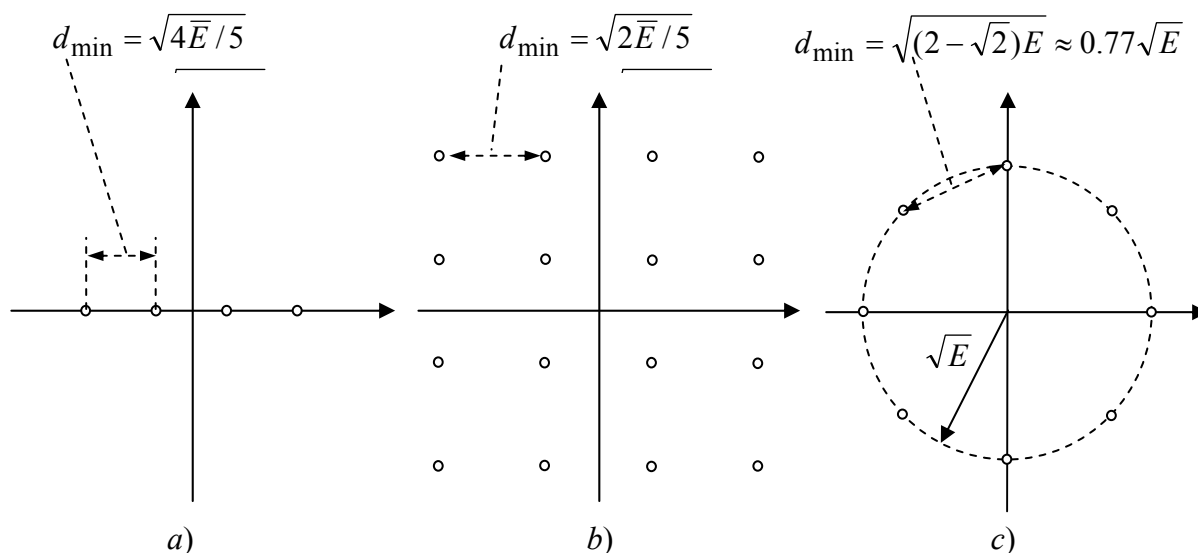


Рис. 2.12. Одно- и двумерные созвездия: 4-АМ (а), 16-КАМ (б), и 8-ФМ (с).

(простейший вариант которой с $M = 2$ – БАМ). Альтернативным наименованием АМ служит *импульсно-кодовая модуляция* (ИКМ)).

В этом случае все сигнальные точки располагаются на одной прямой линии, и при $M > 2$ речь может идти только о задаче «объемной» упаковки. Не составляет труда убедиться, что при заданном $d_{\min}$ оптимальное созвездие, минимизирующее среднюю энергию, образуется в результате равномерного и симметричного расположения сигнальных точек с пространственным разнесением соседних на величину, в точности равной $d_{\min}$ (см. рис. 2.12, а).

При $n_s = 2$ задача нахождения оптимального созвездия с объемной упаковкой становится более трудной и может даже привести к ассиметричным моделям, тогда как сферическая упаковка осуществляется тривиальным образом и реализуется равномерным размещением M точек по окружности радиуса $\sqrt{E}$. Широко используемая в современной цифровой связи M -ичная *квадратурная амплитудная модуляция* (КАМ) служит примером двумерного, симметричного объемно-упакованного созвездия, которое, не являясь оптимальным тео-

ретически, удобно с точки зрения аппаратной реализации (Рис. 2.12, *b*). С другой стороны, традиционная M -ичная *фазовая манипуляция* (М-ФМ) отвечает созвездиям с равномерно расположенными на окружностях точками, оптимальным с точки зрения сферической упаковки (Рис. 2.12, *c*).

Задача оптимального расположения точек в пространстве большей размерности ($n_s > 2$) является чрезвычайно сложной и не имеет до сих пор общего математического решения. Много полезных, но частных результатов содержится в целом ряде книг и статей (см., например, веб-сайт <http://www.research.att.com/~njas/packings/index.html>).

Попытаемся теперь определить верхний предел минимума расстояния в отсутствии предварительного ограничения на размерность сигнального пространства n_s , а также минимальное значение n_s , которое позволит обеспечить найденный предел. Ограничиваясь только случаем сферической упаковки ($E_i = E, i = 1, 2, \dots, M$), вычислим сумму всех M^2 возможных квадратов расстояний, включая и тривиальные (т.е. расстояния для любых сигналов с самим собой). Тогда на основании теоремы косинусов (16) получаем

$$\sum_{i,j=1}^M d^2(\mathbf{s}_i, \mathbf{s}_j) = 2M^2 E - 2E \sum_{i,j=1}^M \rho_{ij}, \quad (2.32)$$

где ρ_{ij} – коэффициент корреляции между i -м и j -м сигналами. Для оценки суммы всех коэффициентов корреляции воспользуемся определением (2.14) для ρ_{ij} , поменяем порядок интегрирования и суммирования и отметим, что двойное суммирование под знаком интеграла осуществляется по отдельным индексам i и j , а значит, его можно заменить произведением двух идентичных сумм:

$$E \sum_{i,j=1}^M \rho_{ij} = \int_0^T \left(\sum_{i,j=1}^M s_i(t) s_j(t) \right) dt = \int_0^T \left(\sum_{i=1}^M s_i(t) \right)^2 dt.$$

Поскольку интеграл от квадрата подынтегрального выражения никогда не будет отрицательным, из (2.32) вытекает, что

$$\sum_{i,j=1}^M d^2(\mathbf{s}_i, \mathbf{s}_j) \leq 2M^2 E.$$

В то же время приведенная выше сумма не меньше, чем $M(M-1)d_{\min}^2$. Учитывая оба указанных неравенства, можно получить следующую верхнюю границу для минимума расстояния

$$d_{\min}^2 \leq \frac{2M}{M-1} E. \quad (2.33)$$

Если существуют M сигналов, достигающие этой верхней границы, то совершенно справедливо будет назвать их оптимальными по критерию минимума расстояния. Для доказательства существования подобного множества сигналов рассмотрим M векторов $\mathbf{u}_i, i = 1, 2, \dots, M$, имеющих попарно нулевое скалярное произведение и единичную длину: $(\mathbf{u}_i, \mathbf{u}_j) = \delta_{ij}, i, j = 1, 2, \dots, M$, где $\delta_{ij} = 0, i \neq j; \delta_{ij} = 1, i = j$ – дельта-функция Кронекера. Такие вектора, называемые *ортонормированными*, существуют в любом векторном пространстве, размерность которого не меньше, чем M . Сформируем теперь M новых векторов $\mathbf{v}_i, i = 1, 2, \dots, M$, каждый из которых образован в виде разности векторов $\mathbf{u}_i$ и их суммы $\mathbf{u} = \sum_{i=1}^M \mathbf{u}_i$, взвешенной коэффициентом $1/M$, т.е. $\mathbf{v}_i = \mathbf{u}_i - \mathbf{u}/M$. Учитывая свойство линейности, скалярное произведение векторов $\mathbf{v}_i$ и $\mathbf{v}_j$ определится как

$$(\mathbf{v}_i, \mathbf{v}_j) = (\mathbf{u}_i, \mathbf{u}_j) - \frac{1}{M}(\mathbf{u}_i, \mathbf{u}) - \frac{1}{M}(\mathbf{u}, \mathbf{u}_j) + \frac{1}{M^2}(\mathbf{u}, \mathbf{u}) = \delta_{ij} - \frac{1}{M}, \quad (2.34)$$

где при вычислении принята во внимание ортогональность векторов $\mathbf{u}_i$. Изменим теперь длину векторов $\mathbf{v}_i$ путем умножения их на $\sqrt{ME/(M-1)}$ и используем полученные вектора в качестве сигнальных:

$$\mathbf{s}_i = \sqrt{\frac{ME}{M-1}} \mathbf{v}_i, i = 1, 2, \dots, M. \quad (2.35)$$

Тогда, согласно теореме косинусов (2.16) и соотношению (2.35), квадрат расстояния между двумя сигналами будет

$$d^2(\mathbf{s}_i, \mathbf{s}_j) = \frac{ME}{M-1} d^2(\mathbf{v}_i, \mathbf{v}_j) = \frac{ME}{M-1} \left[\|\mathbf{v}_i\|^2 + \|\mathbf{v}_j\|^2 - 2(\mathbf{v}_i, \mathbf{v}_j) \right] = \frac{2ME}{M-1}, i \neq j, \quad (2.36)$$

что совпадает с правой частью соотношения (2.33). Следовательно, сигналы, удовлетворяющие границе (2.33) действительно существуют. Более того, расстояние между любой их парой также определяется этой величиной, поэтому приведенные сигналы попадают в категорию *эквидистантных*. Данный класс сигналов широко известен под специальным наименованием *симплексных* сигналов. Непосредственно из их определения вытекает, что

$$\sum_{i=1}^M \mathbf{s}_i = \sqrt{ME/(M-1)} \sum_{i=1}^M \mathbf{v}_i = \sqrt{ME/(M-1)} (\mathbf{u} - \mathbf{u}) = 0,$$

означающее, что в отличие от исходных ортогональных векторов $\mathbf{u}_i$ симплексные сигналы линейно зависимы. Легко убедиться, что размерность пространства $n_s = M-1$, т.е. на единицу меньше числа сигналов, необходима и достаточна для построения M симплексных сигналов.

Свойство эквидистантности симплексных сигналов влечет за собой равенство коэффициентов корреляции ρ_{ij} любой их пары. Оценка величины ρ_{ij} с помощью (2.14), (2.35) и (2.34) приводит к следующему результату

$$\rho_{ij} = \frac{(\mathbf{v}_i, \mathbf{v}_j)}{\|\mathbf{v}_i\| \|\mathbf{v}_j\|} = -\frac{1}{M-1}, i \neq j, i, j = 1, 2, \dots, M,$$

демонстрирующему, что угол между любыми двумя симплексными сигналами один и тот же и больше, чем $\pi/2$. Для простейших множеств из $M = 2, 3, 4$ симплексных сигналов, представленных на рис. 2.13, величина коэффициента корреляции принимает значения, равные -1 (противоположные сигналы), $-1/2$ и $-1/3$ соответственно, что в свою очередь отвечает углам в 180° , 120° и примерно 110° . При $M = 4$ симплексные вектора образуют простейший правильный многогранник (тетраэдр), что объясняет наименование сигналов: «симплекс» на латинском означает «простой».

В любом множестве эквидистантных сигналов $d(\mathbf{s}_i, \mathbf{s}_j) = d_{\min}$ для любой пары различных векторов, так что в (2.35) $n_{\min} = M(M-1)$, что также совпадает с числом слагаемых в (2.34). Подстановка последнего совместно с (2.36) в соотношение (2.35) дает асимптотическую оценку вероятности ошибки, достижимую при использовании симплексных сигналов, которая в то же время, согласно (2.34), служит верхней границей вероятности ошибки

$$P_{e \min} \leq (M-1) Q \left(\sqrt{\frac{ME}{(M-1)N_0}} \right), \frac{2E}{N_0} \gg 1. \quad (2.37)$$

Поскольку симплексные сигналы оптимальны по критерию минимума расстояния, то правая часть последнего выражения одновременно служит асимптотической оценкой минимально возможной вероятности ошибки для случая M сигналов с фиксированными и равными E энергиями.

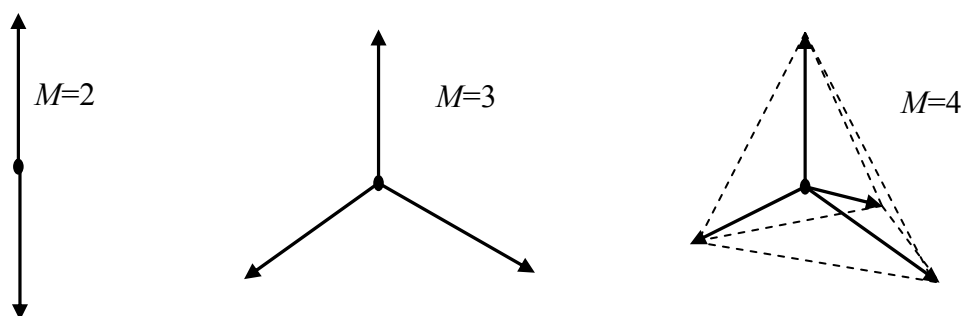


Рис. 2.13. Примеры симплексных сигналов.

Ортогональные сигналы, являющиеся еще одним примером эквидистантных, практически столь же эффективны, как и симплексные, при достаточно большом значении M . Действительно, коэффициент корреляции для ортогональных сигналов равен нулю, а расстояние между любой их парой $d(\mathbf{s}_i, \mathbf{s}_j) = d_{\min} = \sqrt{2E}$. Используя этот факт в (2.35), получим асимптотическую оценку вероятности ошибки для M ортогональных сигналов, которая ограничивает сверху точное значение вероятности ошибки:

$$P_{e \text{ орт}} \leq (M-1) Q\left(\sqrt{\frac{E}{N_0}}\right), \quad \frac{2E}{N_0} \gg 1. \quad (2.38)$$

Сравнение (2.38) с (2.37) демонстрирует, что для выравнивания в обоих случаях вероятностей ошибок необходимо увеличить энергию ортогональных сигналов в $M/(M-1)$ раз по сравнению с симплексными, т.е. энергетические потери γ первых относительно вторых определяются как $\gamma = M/(M-1)$. При $M \gg 1$ эти потери пренебрежимо малы и ортогональные сигналы могут рассматриваться как оптимальные. Например, при $M = 64$ потери $\gamma = 64/63$, что соответствует увеличению энергии ортогональных сигналов по сравнению с симплексными меньше, чем на 0.07 дБ (или 2%). Данное различие, конечно, не имеет практического значения и, если M достаточно велико, ортогональные и симплексные сигналы эквивалентны по помехоустойчивости, так что предпочтение одного ансамбля другому может основываться на реализационных или других соображениях.

Говоря об M -ичной *ортогональной передаче* (в литературе используются также термины *ортогональная модуляция* или *ортогональное кодирование*) вспомним, что максимальное число ортогональных сигналов в точности совпадает с размерностью сигнального пространства, т.е. $M = n_s$. Поэтому в пределах общей полосы W_t и длительности T_t могут быть расположены $W_t T_t$ видео или $2W_t T_t$ полосных ортогональных сигналов. Дополнительное физическое обоснование удвоения числа полосных ортогональных сигналов в сравнении с видеосигналами непосредственно следует из соотношений (2.36) и (2.37). Отметим попутно, что реализация ансамбля M ортогональных сигналов как путем временных или частотных сдвигов между отдельными сигналами, так и путем **кодирования их формы** требует расширения общей полосы частот в M раз. Построение n_s ортогональных сигналов вида (2.36) предполагает добавление к ним еще n_s сигналов, получаемых за счет сдвига фазы несущей частоты на угол $\pi/2$. Эта возможность осуществима на практике только в случае, ко-

гда все сигналы детерминированы или *когерентны*, что означает управление фазами несущей и возможность их использования для идентификации сообщений. В действительности, однако, подобное не осуществимо, поскольку либо сам передатчик, либо канал может нарушить когерентность сигналов, в результате чего их фазы становятся случайными и поэтому не могут использоваться для различения сообщений.

2.4 M-ичная передача данных. Некогерентные сигналы.

Продолжим рассмотрение вопроса M-ичной передачи данных, но теперь в отличие от 2.2 и 2.3 будем полагать, что используемые сигналы не полностью детерминированы. Как уже указывалось ранее, в реальной ситуации высока вероятность того, что передатчик, либо канал не смогут сохранить когерентность полосных сигналов и, значит, на приемной стороне последние будут характеризоваться случайностью фазы. В этом случае значения начальных фаз не могут быть использованы для различения передаваемых сообщений, и отличительный признак сигналов должен строиться не на основе фазового сдвига. Ситуации, отвечающие подобному положению, известны под названием *некогерентного* приема.

Предположим, что закон модуляции i -го полосного сигнала $s_i(t)$ описывается полностью известной комплексной огибающей $\dot{S}_i(t)$ и не изменяющейся во времени случайной начальной фазой φ_i . Тогда данный сигнал представим в виде

$$s_i(t; \varphi_i) = \text{Re}[\dot{S}_i(t; \varphi_i) \exp(j2\pi f_0 t)],$$

где «полная» комплексная огибающая содержит детерминированную составляющую и компоненту, учитывающую случайность начальной фазы: $\dot{S}_i(t; \varphi_i) = \dot{S}_i(t) \exp(j\varphi_i)$.

Введем теперь обобщенную версию скалярного произведения (2.5), которая могла бы использоваться не только в случае реальных сигналов $u(t), v(t)$, но и их комплексных заменителей – аналитических сигналов $\dot{u}(t), \dot{v}(t)$ или комплексных огибающих $\dot{U}(t), \dot{V}(t)$. Модифицированное скалярное произведение может быть определено как

$$(\dot{\mathbf{u}}, \dot{\mathbf{v}}) = \int_0^T \dot{u}(t) \dot{v}^*(t) dt = \int_0^T \dot{U}(t) \dot{V}^*(t) dt = (\dot{\mathbf{U}}, \dot{\mathbf{V}}), \quad (2.39)$$

где комплексное сопряжение используется для сохранения равенства между скалярным произведением вектора с самим собой и квадратом длины вектора (всегда вещественного и не отрицательного), тогда как совпадение скалярного произведения аналитических сигналов и комплексных огибающих следует из определения аналитического сигнала:

$$\dot{s}(t) = \dot{S}(t) \exp(j2\pi f_0 t) = s(t) + js_{\perp}(t), \quad (2.40)$$

где $s_{\perp}(t) = \text{Im}[\dot{S}(t) \exp(j2\pi f_0 t)]$.

В частности, для сигнала $s(t)$ с энергией E согласно соотношениям (2.39) и (2.40) получаем

$$(\dot{\mathbf{S}}, \dot{\mathbf{S}}) = \|\dot{\mathbf{S}}\|^2 = \int_0^T |\dot{S}(t)|^2 dt = \int_0^T S^2(t) dt = \int_0^T s^2(t) dt + \int_0^T s_{\perp}^2(t) dt = 2E, \quad (2.41)$$

поскольку преобразование Гильберта не влияет на амплитудно-частотный спектр и, следовательно, энергии $s(t)$ и $s_{\perp}(t)$ всегда совпадают.

Вычислим квадрат расстояния между комплексными огибающими $\dot{S}_i(t)$ и $\dot{S}_j(t)$:

$$d^2(\dot{\mathbf{S}}_i, \dot{\mathbf{S}}_j) = \|\dot{\mathbf{S}}_i - \dot{\mathbf{S}}_j\|^2 = (\dot{\mathbf{S}}_i - \dot{\mathbf{S}}_j, \dot{\mathbf{S}}_i - \dot{\mathbf{S}}_j) = 2E_i + 2E_j - 4\text{Re}[\dot{\rho}_{ij} \sqrt{E_i E_j}], \quad (2.42)$$

где $\dot{\rho}_{ij}$, как и в (2.16), коэффициент корреляции, но адаптированный к комплекснозначным

сигналам, т.е. комплексным огибающим

$$\dot{\rho}_{ij} = \frac{(\dot{\mathbf{S}}_i, \dot{\mathbf{S}}_j)}{\|\dot{\mathbf{S}}_i\| \|\dot{\mathbf{S}}_j\|} = \frac{1}{2\sqrt{E_i E_j}} \int_0^T \dot{S}_i(t) \dot{S}_j^*(t) dt. \quad (2.43)$$

$$\text{Так как } \int_0^T \dot{S}_i(t) \dot{S}_j^*(t) dt = \int_0^T (s_i(t) + js_{i\perp}(t))(s_j(t) - js_{j\perp}(t)) dt = 2(\mathbf{s}_i, \mathbf{s}_j) + 2j(\mathbf{s}_{i\perp}, \mathbf{s}_j),$$

так что $\text{Re}(\dot{\rho}_{ij}) = \rho_{ij}$, т.е. совпадает с обычным коэффициентом корреляции сигналов $s_i(t)$ и $s_j(t)$, определяемым соотношением (2.14). Учитывая это в (2.42), получаем соотношение, связывающее расстояние между комплексными огибающими и собственно сигналами

$$d^2(\dot{\mathbf{S}}_i, \dot{\mathbf{S}}_j) = 2d^2(\mathbf{s}_i, \mathbf{s}_j). \quad (2.44)$$

Теперь вместо непосредственного вычисления расстояния между сигналами $s_i(t; \varphi_i)$ и $s_j(t; \varphi_j)$ можно воспользоваться соотношением (2.42) и оценить расстояние между комплексными огибающими $\dot{S}_i(t; \varphi_i)$ и $\dot{S}_j(t; \varphi_j)$. Предполагая равенство энергий E всех сигналов, получаем

$$d^2(\dot{\mathbf{S}}_{i\varphi}, \dot{\mathbf{S}}_{j\varphi}) = 4E[1 - \text{Re}\dot{\rho}_{ij}(\varphi)],$$

где дополнительный подстрочный индекс φ подчеркивает соответствие вектора $\dot{\mathbf{S}}_{i\varphi}$ полной комплексной огибающей $\dot{S}_i(t; \varphi_i)$. Тогда, принимая в расчет независимость энергии сигнала E от значения начальной фазы φ_i , и получаем, что

$$\dot{\rho}_{ij}(\varphi) = \frac{(\dot{\mathbf{S}}_{i\varphi}, \dot{\mathbf{S}}_{j\varphi})}{2E} = \frac{1}{2E} \int_0^T \dot{S}_i(t; \varphi_i) \dot{S}_j^*(t; \varphi_j) dt,$$

является коэффициентом корреляции полных комплексных огибающих $\dot{S}_i(t; \varphi_i)$ и $\dot{S}_j(t; \varphi_j)$. Учитывая неизменность значений фаз во времени, последнее соотношение может быть переписано в виде $\dot{\rho}_{ij}(\varphi) = \dot{\rho}_{ij} \exp(\varphi_i - \varphi_j)$, где корреляция $\dot{\rho}_{ij}$, связанная только с детерминированными (т.е. свободными от случайности фазы) комплексными огибающими $\dot{S}_i(t)$ и $\dot{S}_j(t)$ сигналов, представима в виде

$$\dot{\rho}_{ij} = \frac{(\dot{\mathbf{S}}_i, \dot{\mathbf{S}}_j)}{2E} = \frac{1}{2E} \int_0^T \dot{S}_i(t) \dot{S}_j^*(t) dt = |\dot{\rho}_{ij}| \exp(j\varphi_{ij}), \quad (2.45)$$

где $\varphi_{ij} = \arg(\dot{\rho}_{ij})$.

Теперь квадрат расстояния, приведенный выше, будет записан, как

$$d^2(\dot{\mathbf{S}}_{i\varphi}, \dot{\mathbf{S}}_{j\varphi}) = 4E[1 - |\dot{\rho}_{ij}| \cos(\varphi_{ij} + \varphi_i - \varphi_j)]. \quad (2.46)$$

Единственная трудность, возникающая при его вычислении, заключается в его зависимости от неизвестных фаз φ_i, φ_j . Вследствие этого, существует множество расстояний для заданных детерминированных законов модуляции $\dot{S}_i(t)$ и $\dot{S}_j(t)$, отвечающих случайной разности фаз $\varphi_i - \varphi_j$. При значительном уровне модуля коэффициента корреляции $|\dot{\rho}_{ij}|$ всегда присутствует вероятность, что при неудачной комбинации фаз (т.е. при малости $\varphi_{ij} + \varphi_i - \varphi_j$) расстояние (2.46) окажется малым. Для исключения подобной ситуации необходимо, чтобы модуль коэффициента корреляции был минимально возможным, а наилучшее множество

сигналов должно удовлетворять условию

$$\dot{\rho}_{ij} = \frac{(\dot{\mathbf{S}}_i, \dot{\mathbf{S}}_j)}{2E} = 0, i \neq j, i, j = 1, 2, \dots, M. \quad (2.47)$$

Вновь, как и в 2.3, приходим к необходимости использования ортогональных сигналов. Однако теперь условие ортогональности является более обязывающим, обуславливая ортогональность комплексных огибающих, или, другими словами, законов модуляции, сигналов, а не только самих сигналов. Вследствие этого полосные сигналы сохраняют ортогональность при любых сочетаниях фаз, поскольку $\dot{\rho}_{ij} = 0$ влечет и $\dot{\rho}_{ij}(\varphi) = 0$. С другой стороны, условие (2.47) исключает возможность обеспечения ортогональности за счет квадратурного сдвига фазы несущей, что было приемлемым в случае когерентных полосных сигналов. Прямым следствием этого является сокращение вдвое размерности сигнального пространства n_s с величины $2W_t T_t$, как это было в случае видеосигналов, до значения $W_t T_t$ при заданной полосе W_t . Данный факт может быть объяснен и несколько иным образом: ортогональность должна обеспечиваться для комплексных сигналов, т.е. в комплексном векторном пространстве. Каждый комплексный вектор представим в виде двух, отвечающих вещественной и мнимой частям, и, следовательно, требует для своего представления двумерного векторного пространства, т.е. плоскости. Все эти плоскости должны быть ортогональны друг другу. Поскольку максимально достижимая размерность пространства составляет величину $2W_t T_t$, то в нем можно расположить только $W_t T_t$ ортогональных плоскостей, а, значит, и $W_t T_t$ ортогональных комплексных сигналов.

Следует упомянуть, что в отличие от ситуации, которая имела место для M детерминированных сигналов, при некогерентном приеме ортогональные сигналы являются строго оптимальными вне зависимости от их числа. Например, при некогерентном приеме оптимальной является пара ортогональных сигналов, тогда как пара противоположных сигналов не удовлетворяет данному утверждению, поскольку случайность фазы исключает возможность их различения.

Еще один момент, о котором следует упомянуть, касается выяснения оптимальной стратегии решения при некогерентном приеме. Наблюдение $y(t)$, рассматриваемое как полосный сигнал, может быть выражено в терминах комплексной огибающей или закона модуляции $\dot{Y}(t)$: $y(t) = \text{Re}[\dot{Y}(t) \exp(j2\pi f_0 t)]$. Очевидно, что $\dot{Y}(t)$ является случайным процессом. При решении, какой из M возможных сигналов был принят, необходимо сравнить расстояния между наблюдением и сигналами, что согласно соотношению (2.44), эквивалентно сравнению квадратов расстояний между соответствующими комплексными огибающими: $d^2(\dot{\mathbf{S}}_{i\varphi}, \dot{\mathbf{Y}}) = \|\dot{\mathbf{Y}}\|^2 + 2E - 2\text{Re}[\dot{z}_i(\varphi_i)]$, где $\dot{z}_i(\varphi_i)$ – значение корреляции (скалярного произведения) комплексных огибающих наблюдения $y(t)$ и i -го сигнала со случайной фазой $s_i(t; \varphi_i)$. Для исключения зависимости от случайной фазы логично выбрать для сравнения только минимальное по всем возможным значениям φ_i расстояние для каждого i . Опуская детали, воспроизводящие почти дословно приведенные выше рассуждения, окончательное решающее правило выражается в терминах модуля корреляции

$$Z_i = \left| \int_0^T \dot{Y}(t) \dot{S}_i^*(t) dt \right|, i = 1, 2, \dots, M, \quad (2.48)$$

означающее, что j -й сигнал считается принятым, если модуль корреляции между его полнотой известной комплексной огибающей и аналогичной характеристикой наблюдения принимает максимальное значение. Другими словами, используется старое и неоднократно используемое правило: выбирается такой сигнал, закон модуляции которого в наибольшей

степени подобен закону модуляции наблюдения.

2.4.1 Некогерентные двоичные сигналы

При анализе когерентного приема бинарных сигналов и разумном предположении об их равных априорных вероятностях передачи ($P(s_0) = P(s_1)$) канал являлся симметричным, т.е. вероятности перепутывания символов были одинаковы: $P(s_0/s_1) = P(s_1/s_0)$. Эту ситуацию иллюстрирует рис. 2.14, где $W(Z/0)$ и $W(Z/1)$ – мгновенные распределения сигналов s_0 и s_1 .

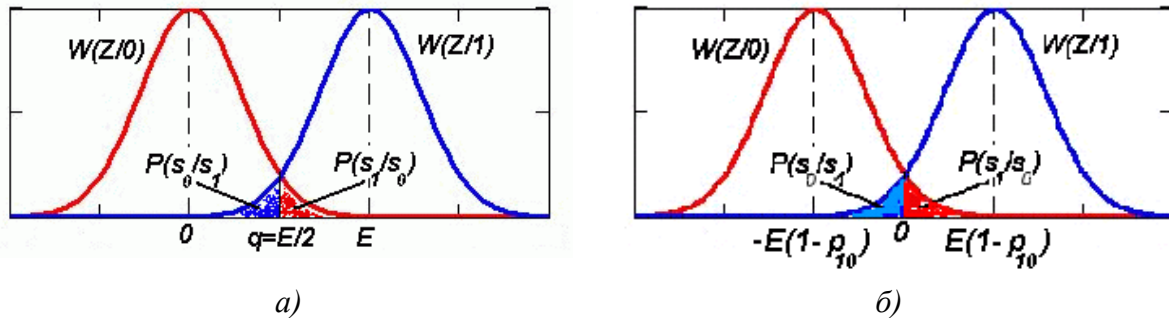


Рис. 2.14. Плотности распределения сигналов на входе решающей схемы при когерентном приеме: а) – БМ; б) – БЧМ и БФМ

Вероятность ошибки при этом определялась, как $P_E = P(s_0/s_1) = P(s_1/s_0)$.

Амплитудно-манипулированные сигналы

Для некогерентного приема огибающая сигналов БМ представлена на рис. 2.15, а ее распределения – на рис. 2.16.

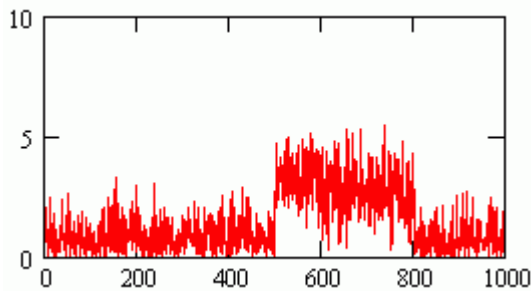


Рис. 2.15. Огибающая сигнала БМ.

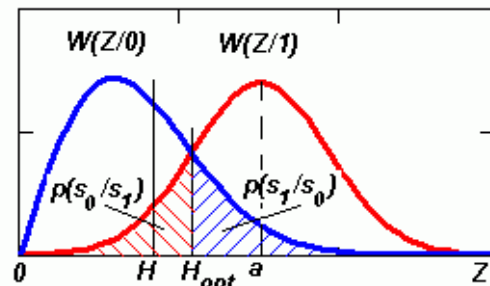


Рис. 2.16. Распределения огибающей БМ.

Распределение огибающей $U_{сш}$ подчиняется обобщенному закону Релея (Релея-Райса)

$$W(U_{сш}) = \frac{U_{сш}}{\sigma^2} e^{-\frac{U_{сш}^2 + S_0^2}{2\sigma^2}} I_0\left(\frac{S_0 U_{сш}}{\sigma^2}\right),$$

где S_0 – амплитуда полезного сигнала, $I_0(\cdot)$ – модифицированная функция Бесселя нулевого порядка, и который переходит в закон Релея при отсутствии сигнала:

$$W(U_{ш}) = \frac{U_{ш}}{\sigma^2} e^{-U_{ш}^2 / 2\sigma^2}.$$

Переходя к нормированным величинам $a = S_0/\sigma$ и $Z = U_{сш}/\sigma$, запишем эти распределения в виде

$$W(Z/1) = Ze^{-(Z^2 + a^2)/2} I_0(aZ) \quad \text{и} \quad (2.49)$$

$$W(Z/0) = Ze^{-Z^2/2} \quad (2.50)$$

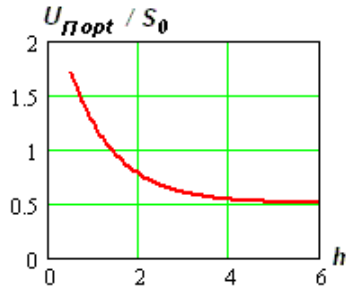
Теперь, даже при равенстве априорных вероятностей $P(s_0) = P(s_1)$, канал несимметричен и вероятности перепутывания символов определятся, как

$$P(s_0 / s_1) = \int_0^H W(Z/1) dZ \quad \text{и} \quad P(s_1 / s_0) = \int_H^\infty W(Z/0) dZ,$$

где H – нормированный порог: $H = U_{nop} / \sigma$.

Оптимальное значение порога, минимизирующее вероятность ошибки $P_E = 0,5[P(s_0/s_1) + P(s_1/s_0)]$, как видно из рис. 2.9, находится как абсцисса точки пересечения кривых $W(Z/0)$ и $W(Z/1)$. Решая уравнение $W(Z/0)|_{Z=H_{opt}} = W(Z/1)|_{Z=H_{opt}}$ относительно Z , получим $I_0(aH_{opt}) = \exp(a^2/2)$. Решение этого трансцендентного уравнения представлено графиком на рис. 2.17,

где $h = \sqrt{\frac{E}{N_0}}$ – ОСШ.



При достаточно большом ($h \geq 3$) отношении сигнал/шум величина оптимального порога $H_{opt} = U_{nop} / S_0$ стремится к 0,5, т.е. $H_{opt} = a/2 = h/\sqrt{2}$. При $h \geq 3$ закон Релея-Райса достаточно хорошо аппроксимируется нормальным законом:

$$W(Z/1) \approx \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(Z-a)^2}{2}\right], \quad a \geq 4$$

Рис.2.17. Зависимость оптимального порога от отношения сигнал/шум

Используем эту аппроксимацию для вычисления

условной вероятности перепутывания «1» с «0» (пропуск сигнала):

$$P(s_0 / s_1) = 1 - \frac{1}{\sqrt{2\pi}} \int_{H_{opt}}^\infty \exp\left[-\frac{(Z-a)^2}{2}\right] dZ =$$

$$1 - \frac{1}{\sqrt{2\pi}} \int_{H_{opt}-a}^\infty \exp(-y^2 / 2) dy = 1 - \quad (2.51)$$

$$\frac{1}{\sqrt{2\pi}} \int_{-H_{opt}}^\infty \exp(-y^2 / 2) dy = 0,5 - \hat{Q}(H_{opt}) = Q(H_{opt})$$

Здесь использовано $\int_{-x}^\infty W(z) dz = \int_{-x}^0 W(z) dz + \int_0^\infty W(z) dz = \frac{1}{2} + \int_{-x}^0 W(z) dz$

Вероятность ложной тревоги $P(s_1 / s_0)$ определится, как

$$P(s_1 / s_0) = \int_{H_{opt}}^\infty z e^{-z^2/2} dz = \int_{H_{opt}}^\infty e^{-z^2/2} d\frac{z^2}{2} = e^{-H_{opt}^2/2} \quad (2.52)$$

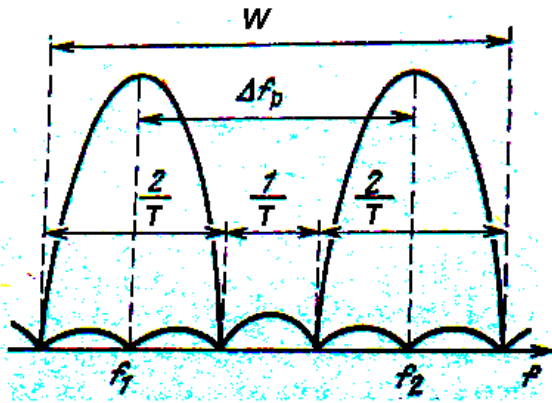
Полная вероятность ошибки

$$P_E = 0,5 \{P(s_0 / s_1) + P(s_1 / s_0)\} = 0,5 \left\{ e^{-H_{opt}^2/2} + Q(H_{opt}) \right\} \quad (2.53)$$

Т.к. $H_{opt} \approx a/2 = h/\sqrt{2}$ при больших h , то

$$P_E = 0,5 \left\{ e^{-h^2/4} + Q(h/\sqrt{2}) \right\}, \quad h \geq 3. \quad (2.54)$$

Частотно-манипулированные сигналы



Спектр частотно-манипулированных (БЧМ) сигналов представлен на рис. 2.18. Полный сигнал занимает полосу W , главный лепесток каждого из них имеет ширину $1/T$, частоты разнесены на величину Δf_p .

Оптимальный в смысле экономии частотного ресурса разнос частот $\Delta f_{opt} = 0,75/T$. Ниже этого значения различие сигналов ухудшается, его увеличение не приводит к росту помехоустойчивости.

Рис. 2.18. Спектр сигналов БЧМ

Минимально допустимая ширина спектра сигналов БЧМ равна $W = \Delta f_p + 1/T \approx 2/T$, а аналитическая запись такого набора имеет вид

$$s(t) = \begin{cases} s_1(t) = S_0 \cos(\omega_1 t + \varphi_1) \\ s_0(t) = S_0 \cos(\omega_2 t + \varphi_2), \text{ где } 0 \leq t \leq T \end{cases} \quad (2.55)$$

Некогерентный приемник для этих сигналов можно строить как с частотным дискриминатором (ЧД), так и с двумя полосовыми фильтрами. Приемник с ЧД, вследствие нелинейности последнего, уступает линейному приемнику с полосовой фильтрацией.

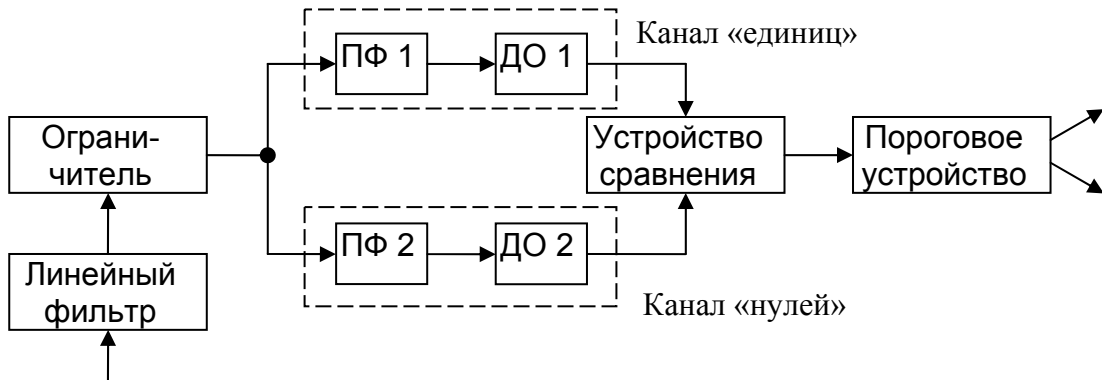


Рис. 2.19. Приемник ЧМ-сигналов с двумя полосовыми фильтрами

Схема приемника, приведенная на рис. 2.19, симметрична относительно “1” и “0”, поэтому $P(s_1/s_0) = P(s_0/s_1)$. Найдем вероятность ошибки при его использовании. Пусть передается сигнал s_1 (“1”). Тогда $U_{сш}$ – выход канала «единиц», $U_{ш}$ – выход канала «нулей». Ошибка происходит, когда $U_{ш} \geq U_{сш}$. Вероятность этого события

$$P(U_{ш} \geq U_{сш}) = \int_{U_{сш}}^{\infty} W(U_{ш}) dU_{ш} \quad (2.56)$$

$U_{сш}$ – случайная величина, поэтому необходимо произвести усреднение по всем возможным значениям $U_{сш}$:

$$P(s_0/s_1) = \int_0^{\infty} W(U_{ш}) P(U_{сш} \geq U_{ш}) dU_{ш} \text{ или}$$

$$P(s_0 / s_1) = \int_0^\infty \int_{U_{\tilde{N}\emptyset}}^\infty W(U_{\tilde{N}\emptyset}) W(U_\emptyset) dU_\emptyset dU_{\tilde{N}\emptyset}, \quad (2.57)$$

где

$W(U_{C\text{ш}})$ – распределение Релея-Райса,

$W(U_{\text{ш}})$ – распределение Релея. Используя $v = U_{C\text{ш}} / \sigma$, $z = U_{\text{ш}} / \sigma$, $a = S_0 / \sigma$, получим:

$$P(s_0 / s_1) = \int_0^\infty \int_{U_{\tilde{N}\emptyset}}^\infty \frac{U_{\tilde{N}\emptyset}}{\sigma^2} \exp\left(-\frac{U_{\tilde{N}\emptyset}^2 + S_0^2}{2\sigma^2}\right) I_0\left(\frac{U_{\tilde{N}\emptyset} S_0}{\sigma^2}\right) \frac{U_\emptyset}{\sigma^2} \exp\left(-\frac{U_\emptyset^2}{2\sigma^2}\right) dU_\emptyset dU_{\tilde{N}\emptyset} =$$

$$\int_0^\infty \int_{U_{\tilde{N}\emptyset}/\sigma}^\infty v \exp\left(-\frac{v^2 + a^2}{2}\right) I_0(av) z \exp\left(-\frac{z^2}{2}\right) dv dz; \quad (2.58)$$

Двойной интеграл легко разделяется, причем [6]:

$$\int_0^\infty v \exp(-v^2) I_0(av) dv = \frac{1}{2} \exp\left(\frac{a^2}{4}\right) \text{ и окончательно}$$

$$P(s_0 / s_1) = 0,5 \exp\left(-\frac{a^2}{4}\right) = 0,5 \exp\left(-\frac{h^2}{2}\right)$$

Итак, вероятность ошибки при некогерентном приеме сигналов БЧМ

$$P_E = 0,5 \exp\left(-\frac{h^2}{2}\right) \quad (2.59)$$

2.4.2 Сравнение когерентного и некогерентного способов приема бинарных сигналов

Определим проигрыш некогерентного (НКГ) приема когерентному (КГ) в предположении оптимальной полосы приемника W : $WT \approx 1$, T – длительность сигнала.

Отношение С/Ш на входе НКГ приемника при его оптимальной полосе

$$h^2 = \left(\frac{P_C}{P_\emptyset}\right)_{A\emptyset} = \frac{S_0^2}{2N_0W} = \frac{S_0^2T}{2N_0} = \frac{E}{N_0} = h_0^2, \text{ что совпадает с когерентным случаем.}$$

Для отличия обозначим его как $h_{\text{нкг}}$.

Функция $Q(x)$ допускает разложение в ряд:

$$Q(x) = \frac{\exp(-x^2/2)}{x\sqrt{2\pi}} \left(1 - \frac{1}{x^2} + \frac{1}{x^4} - \dots\right), \quad (2.60)$$

который быстро сходится при $x > 1$:

$$Q(x) \approx \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}, \quad x \geq 3.$$

А. Амплитудная манипуляция

Используем разложение (2.60) для записи вероятности ошибок при амплитудной манипуляции

$$\begin{cases} P_{E\hat{E}\hat{A}} = \{Q(h_0 / \sqrt{2})\} \approx \frac{1}{h_0 \sqrt{\pi}} e^{-h_0^2/4} \\ P_{\hat{A}\hat{I}\hat{E}\hat{A}} \approx 0,5 e^{-h_{0i\hat{e}\hat{a}}^2/4} \end{cases}$$

Равные энергетические затраты

Сравниваем методы приема при равных затратах энергии ($h_{0\text{нкз}} = h_0$), т.е. по величине вероятности ошибок:

$$\frac{P_{E\hat{I}\hat{E}\hat{A}}}{P_{E\hat{E}\hat{A}}} = \frac{\sqrt{\pi} h_0}{2}.$$

Проигрыш НКГ приема сигналов БАМ когерентному пропорционален h_0 .

Равные ошибки

Найдем энергетический проигрыш при равенстве вероятностей ошибок $P_{E\text{НКГ}} = P_E$ КГ, что приводит к уравнению

$$\frac{1}{h_0 \sqrt{\pi}} e^{-h_0^2/4} = 0,5 e^{-h_{0i\hat{e}\hat{a}}^2/4}, \text{ из которого следует, что}$$

$$\frac{h_{0i\hat{e}\hat{a}}^2}{h_0^2} = 1 + \frac{4}{h_0^2} \ln \left(\frac{h_0 \sqrt{\pi}}{2} \right)$$

Расчеты показывают, что для обеспечения вероятности ошибки от 10^{-3} до 10^{-6} НКГ способ приема требует увеличения затрат энергии на (30 – 15)%. При увеличении отношения сигнал/шум разница между когерентным (оптимальным) и некогерентным приемом нивелируется.

Б. частотная манипуляция

Ошибки при КГ приеме ЧМн-сигналов: $P_E = Q(h_0)$.

Используем разложение в ряд при $h_0 > 3$:

$$P_{E\hat{E}\hat{A}} = \frac{1}{h_0 \sqrt{2\pi}} \exp \left(-\frac{h_0^2}{2} \right)$$

Равные энергетические затраты

$h_{0\text{нкз}} = h_0$. Из (2.58) и (2.59) следует

$$\frac{P_{E\hat{I}\hat{E}\hat{A}}}{P_{E\hat{I}\hat{I}\hat{A}}} = \sqrt{\frac{\pi}{2}} h_0 \approx 1,26 h_0$$

Равные ошибки

$$\frac{h_{0i\hat{e}\hat{a}}^2}{h_0^2} = \frac{1 + 2 \ln(1,26 h_0)}{h_0^2}$$

Как и в случае БАМ, при $P_E = 10^{-3}..10^{-6}$ энергетический проигрыш некогерентного метода составляет (30 - 15)%, т.е. всего 1 – 0,5 дБ и уменьшается с ростом h_0 .

2.5 Примеры множеств ортогональных сигналов.

В данном параграфе, как и ранее, не будет рассматриваться возможность удвоения числа ортогональных сигналов за счет квадратурного сдвига несущей частоты, что всегда имеет место в когерентных полосных системах, а будет уделено основное внимание зависимости между M и эквивалентным частотно-временным ресурсом $W_t T_t$ системы. Первона-

чально продемонстрируем возможность построения простейших множеств ортогональных сигналов за счет дробления доступного ресурса.

Кодирование путем временного сдвига

Очевидно, что скалярное произведение любых двух не перекрывающихся во времени сигналов равно нулю. Рассмотрим M сигналов, приведенных на рис. 2.20, *a*, которые совместно занимают весь временной интервал T_t . При длительности сигнала не более, чем

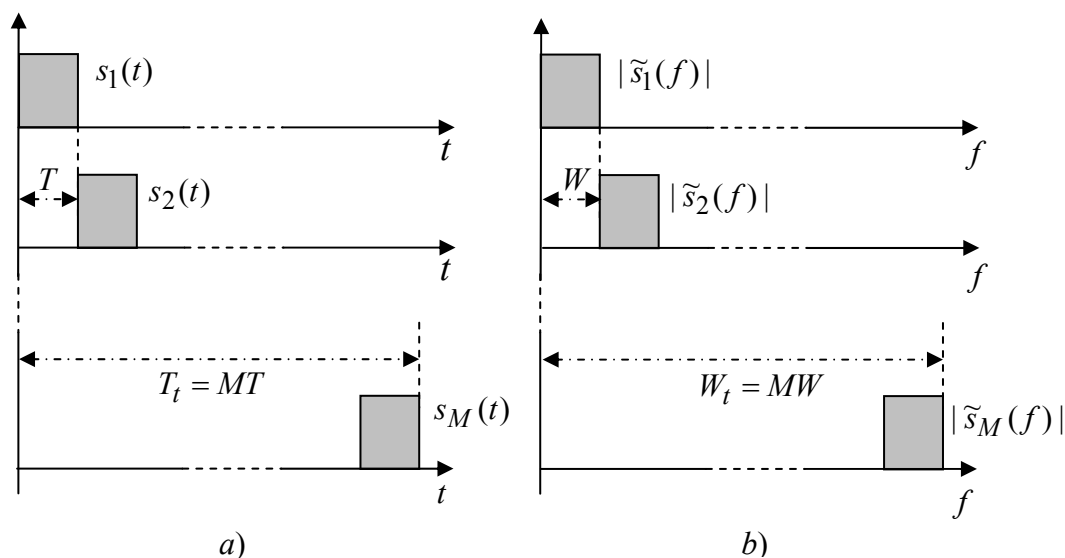


Рис. 2.20. Ортогональные по времени (*a*) и частоте (*b*) сигналы.

$T = T_t / M$, и временном сдвиге между соседними сигналами не менее, чем длительность сигнала, подобное кодирование путем временного сдвига образует семейство ортогональных сигналов. Оценка полосы W , занимаемая каждым из сигналов, равна обратной величине от их длительности, и все сигналы, не нарушая условия ортогональности, занимают одну и ту же полосу $W = W_t$. Следовательно, максимальное число подобных ортогональных сигналов, которые могут располагаться в пределах заданного частотно-временного ресурса W_t, T_t , составляет величину $M = T_t / T = W_t T_t$, т.е., как можно было и предвидеть, равняется размерности сигнального пространства $n_s = W_t T_t$. Необходимость большего числа сигналов $M \gg 1$ требует большой величины произведения $W_t T_t = M$, которое, по-видимому, может быть достигнуто путем расширения спектра. Однако для любого конкретного сигнала частотно-временное произведение $WT = W_t T = W_t T_t / M = 1$, так что данные сигналы не относятся к типу с распределенным спектром.

Пусть общий частотно-временной ресурс представлен в виде прямоугольника на координатной плоскости время-частота со сторонами, равными соответственно T_t и W_t . Тогда кодирование путем временного сдвига означает лишь разрезание этого ресурса на M вертикальных полосок, каждая из которых приписывается некоторому конкретному сигналу (см. рис. 2.21, *a*). Ортогональность при передаче обеспечивается строгим распределением временного ресурса между сигналами, каждый из которых занимает весь отведенный частотный ресурс.

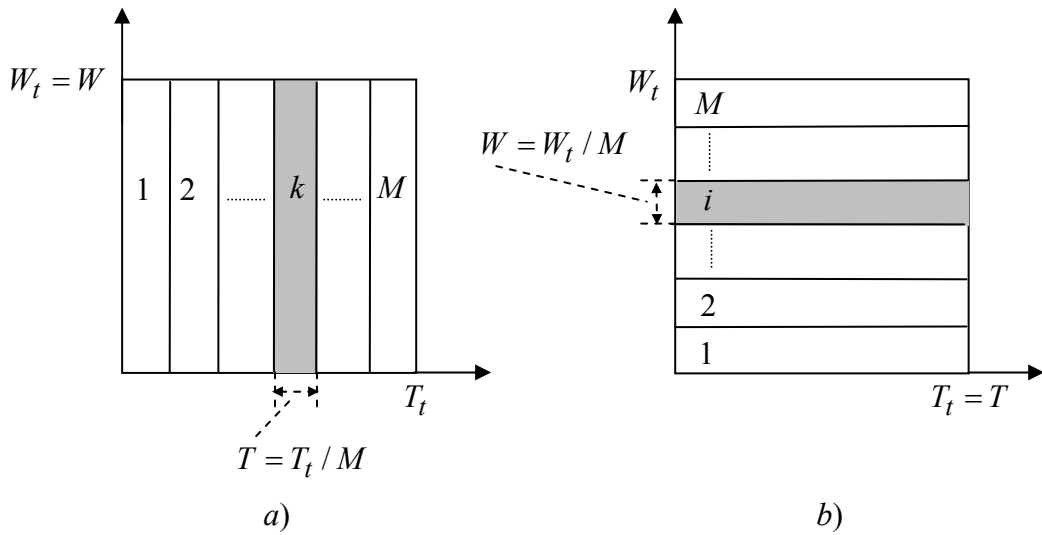


Рис. 2.21. Распределение ресурса при временном (а) и частотном (b) ортогональном кодировании.

Представленная выше схема ортогональной сигнализации может показаться привлекательной с точки зрения практической реализации благодаря ее явной простоте. Ее слабость, однако, также заметна и ее следует учитывать. Во-первых, необходима точная синхронизация, поскольку любое потенциальное изменение временной позиции сигнала может служить причиной нарушения ортогональности и требует надежного сохранения границ между сигналами. Достижение последнего сокращает возможное число сигналов по сравнению с теоретическим максимумом, т.е. ухудшает спектральную эффективность. Другим недостатком является значение пик-фактора ν , который определяется отношением пиковой мощности к средней. Поскольку индивидуальный сигнал занимает только M -ю часть доступного временного ресурса, то средняя мощность в M раз меньше пиковой и $\nu = M \gg 1$. Между тем при проектировании усилителя мощности передатчика малое значение ν является необходимым: чем ближе его значение к единице, тем более мягкие требования к линейности усилителя и тем лучше его мощностные характеристики.

Кодирование путем частотного сдвига

Другим непосредственным способом обеспечения ортогональности служит *кодирование путем частотного сдвига*. На основании дуальности времени и частоты или теоремы Парсеваля скалярное произведение сигналов $u(t)$, $v(t)$ и их спектров $\tilde{u}(f)$, $\tilde{v}(f)$ совпадает:

$$(\mathbf{u}, \mathbf{v}) = \int_{-\infty}^{\infty} u(t) v(t) dt = \int_{-\infty}^{\infty} \tilde{u}(f) \tilde{v}^*(f) df = (\tilde{\mathbf{u}}, \tilde{\mathbf{v}}), \quad (2.61)$$

что позволяет перевести ранее представленный метод в частотную область (см. Рис. 2.13, b). При полном перекрытии сигналов во времени ($T = T_t$) каждый из них занимает полосу не менее чем $W = 1/T_t$. Тогда максимальное число ортогональных сигналов, образованных сдвигом спектра, снова будет $M = W_t / W = W_t T_t = n_s$. Как и в предыдущем случае, общий ресурс снова нарезается на полоски, однако иным образом. Полоски располагаются горизонтально, означая, что любой сигнал использует весь временной ресурс T_t и только M -ю часть общего частотного ресурса W_t (рис. 2.14, b). Очевидно, что каждый индивидуальный сигнал снова не является сигналом с распределенным спектром, поскольку отвечающее ему частотно-временное произведение $WT = (W_t / M)T = 1$.

В отличие от кодирования временным сдвигом, пик-фактор ортогональных сигналов

данного вида $v = 1$ и ошибки в синхронизации не играют столь важную роль, поскольку ортогональность обеспечивается за счет отсутствия перекрытия в частотной области. Вместо этого деструктивную роль в некоторых случаях играет дрейф спектра (из-за смещения Доплера). Несмотря на это, данный способ передачи чрезвычайно популярен и примером его прямой реализацией служит обычная M -ичная частотная модуляция.

Ортогональное кодирование с распределением спектра

Дробление общего частотно-временного ресурса, присущего двум ранее рассмотренным методам построения ортогональных сигналов, в некоторых случаях может оказаться предпочтительным с точки зрения аппаратной реализации. Однако при увеличении числа сигналов M подобные основания становятся сомнительными, поскольку, как уже упоминалось, кодирование с помощью временного сдвига требует значительного пик-фактора, а кодирование с помощью частотного сдвига приводит к оптимальной обработке, состоящей в использовании значительного набора параллельных частотных фильтров.

При таких обстоятельствах метод построения ортогональных сигналов путем распределения спектра может оказаться весьма конкурентоспособным, поскольку предполагает совместное использование всеми сигналами общего частотно-временного ресурса без распределения или нарезания последнего. Рассмотрим простой пример воплощения указанной идеи для дискретных бинарных фазоманипулированных сигналов. Образует любой из M сигналов в виде последовательности N примыкающих друг другу элементарных импульсов (или чипов), каждый из которых обладает прямоугольной формой и длительностью Δ . Пусть полярность чипов сигнала под номером k изменяется в соответствии с *кодовой последовательностью* (или просто *кодом*) бинарных символов $a_{k,i} = \pm 1$, где $k = 1, 2, \dots, M$, а второй подстрочный символ отвечает номеру чипа (дискретному времени) $i = 0, 1, \dots, N-1$. Тогда версия подобного сигнала на видеочастоте может быть записана в виде

$$s_k(t) = \sum_{i=0}^{N-1} a_{k,i} s_0(t - i\Delta), \quad (2.62)$$

где $s_0(t)$ символизирует прямоугольный импульс длительности Δ .

Вычислим теперь скалярное произведение или корреляцию (2.5) k -го и l -го сигналов. После изменения порядка суммирования и интегрирования получим

$$(s_k, s_l) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} a_{k,i} a_{l,j} \int_0^T s_0(t - i\Delta) s_0(t - j\Delta) dt. \quad (2.63)$$

В последнем соотношении интеграл представляет собой скалярное произведение двух сдвинутых во времени друг относительно друга чипов на величину, равную $(i - j)\Delta$. При $i \neq j$ интеграл равен нулю, поскольку импульсы не перекрываются во времени. Таким образом

$$\int_0^T s_0(t - i\Delta) s_0(t - j\Delta) dt = E_0 \delta_{ij},$$

где E_0 – энергия элементарного импульса. Использование полученного выражения в (2.63) дает

$$(s_k, s_l) = E_0 \sum_{i=0}^{N-1} a_{k,i} a_{l,i} = E_0 (\mathbf{a}_k, \mathbf{a}_l). \quad (2.64)$$

Последнее соотношение связывает скалярное произведение сигналов (2.62) со скалярным произведением N -мерных векторов, отвечающих кодовым последовательностям $\mathbf{a}_k = (a_{k,0}, a_{k,1}, \dots, a_{k,N-1})$. Очевидно, что M ортогональных кодовых последовательностей автоматически формируют M ортогональных сигналов типа (2.62). При $M \leq N$ существует множество способов конструирования подобных последовательностей, поскольку рас-

рассматриваемая ситуация отвечает задаче нахождения $M \leq N$ ортогональных векторов размерности N . В нашем примере упомянутые вектора являются бинарными, т.е. их компоненты принимают значения из множества ± 1 . При $M = N$ ортогональные бинарные вектора, рассматриваемые в качестве строк, образуют квадратную матрицу, называемую матрицей Адамара. Не составляет труда убедиться, что существуют лишь матрицы Адамара размерности, кратной четырем, т.е. $M \equiv 0 \pmod{4}$, где используемое сравнение $a \equiv b \pmod{c}$ означает равенство остатков от деления целых чисел a, b на целое c . До сих пор не найдено доказательство достаточности этого необходимого условия.

Известен целый ряд алгоритмов построения матриц Адамара определенной (не произвольной) длины. Наиболее популярным из них является правило Сильвестра, позволяющее рекурсивно строить матрицы удвоенного размера. Для пояснения данного алгоритма предположим, что так или иначе была найдена матрица Адамара $\mathbf{H}_M$ размерности M . Тогда матрица Адамара $\mathbf{H}_{2M}$ удвоенного размера может быть построена путем четырехкратного повторения $\mathbf{H}_M$, взятой в качестве блоков $\mathbf{H}_{2M}$, один из которых с противоположным знаком:

$$\mathbf{H}_{2M} = \begin{bmatrix} \mathbf{H}_M & \mathbf{H}_M \\ \mathbf{H}_M & -\mathbf{H}_M \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \otimes \mathbf{H}_M, \quad (2.65)$$

где второе равенство выражает правило кронекеровского произведения матриц. Ортогональность строк $\mathbf{H}_{2M}$ очевидна: две строки, номера которых отличаются на любое целое, но не M , обладают нулевым скалярным произведением, поскольку их две M -элементные половины ортогональны просто потому, что они представляют собой ортогональные строки матрицы $\mathbf{H}_M$. В остальных случаях первые M компонент строк совпадают, тогда как остальные противоположны, что снова обеспечивает ортогональность.

Применение алгоритма Сильвестра начинается с матрицы $\mathbf{H}_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$, которая является простейшей матрицей Адамара. Следующий шаг состоит в построении матрицы $\mathbf{H}_4$, в которой для краткости обозначим символами «+» и «-» значения элементов «+1» и «-1». Затем от $\mathbf{H}_4$ перейдем к $\mathbf{H}_8$ и т.д.:

$$\mathbf{H}_4 = \begin{bmatrix} \mathbf{H}_2 & \mathbf{H}_2 \\ \mathbf{H}_2 & -\mathbf{H}_2 \end{bmatrix} = \begin{bmatrix} + & + & + & + \\ + & - & + & - \\ + & + & - & - \\ + & - & - & + \end{bmatrix}$$

$$\mathbf{H}_8 = \begin{bmatrix} \mathbf{H}_4 & \mathbf{H}_4 \\ \mathbf{H}_4 & -\mathbf{H}_4 \end{bmatrix} = \begin{bmatrix} + & + & + & + & + & + & + & + \\ + & - & + & - & + & - & + & - \\ + & + & - & - & + & + & - & - \\ + & - & - & + & + & - & - & + \\ + & + & + & + & - & - & - & - \\ + & - & + & - & - & + & - & + \\ + & + & - & - & - & - & + & + \\ + & - & - & + & - & + & + & - \end{bmatrix}$$

Таким образом, согласно данному алгоритму может быть построена любая матрица Адамара размерности $M = 2^m$ (2, 4, 8, 16, 32, ...). Строки матрицы Адамара, построенной подобным образом, известны как *функции Уолша*.

На рис. 2.21 представлены низкочастотные ортогональные бинарные фазоманипулированные сигналы (2.62) – функции Уолша, построенные на основе матрицы Адамара H_8 .

Рис. 2.21 иллюстрирует тот факт, что в рамках данного метода построения ортого-

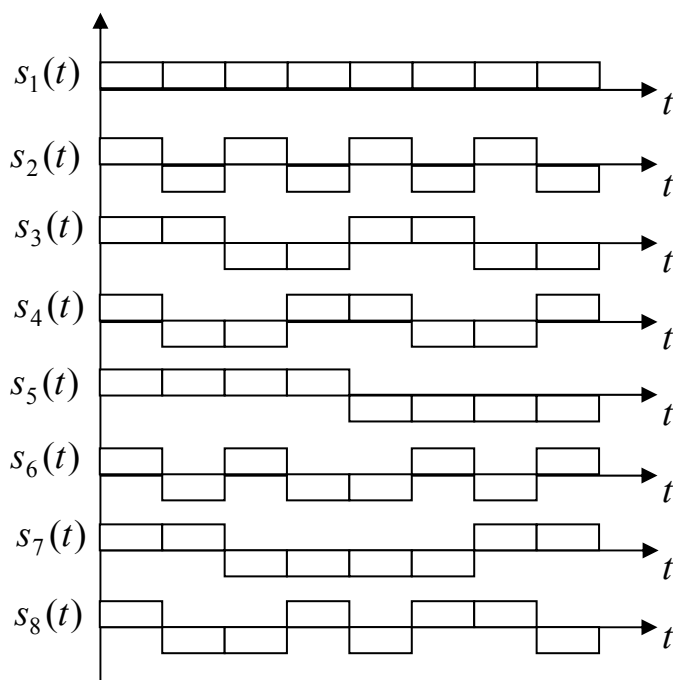


Рис. 2.21. Функции Уолша на основе H_8 .

нальных сигналов, не происходит деления общего ресурса: все сигналы располагаются в нем, полностью перекрываясь как во временной, так и в частотной области. Действительно, полоса, занимаемая каждым сигналом, может быть оценена как $W = 1/\Delta$, тогда как длительность составляет величину $T = M\Delta$, откуда $WT = M = W_t T_t$. Ортогональность же в данном методе достигается не путем деления временного интервала или полосы, а за счет соответствующего выбора закона модуляции сигнала.

Анализируя достоинства ортогональности за счет распределения спектра, можно отметить, что методы формирования и обработки сигналов (2.62) хорошо сопрягаются с современными средствами цифровой микроэлектроники. Еще одной особенностью служит автоматическое приобретение тех преимуществ технологии распределенного спектра, которые незаметны в рамках классической структуры приема, но многочисленны и очень ценны на практике. Это объясняет значительное распространение ортогональной сигнализации данного типа в современных телекоммуникационных системах (cdmaOne, UMTS, cdma2000).

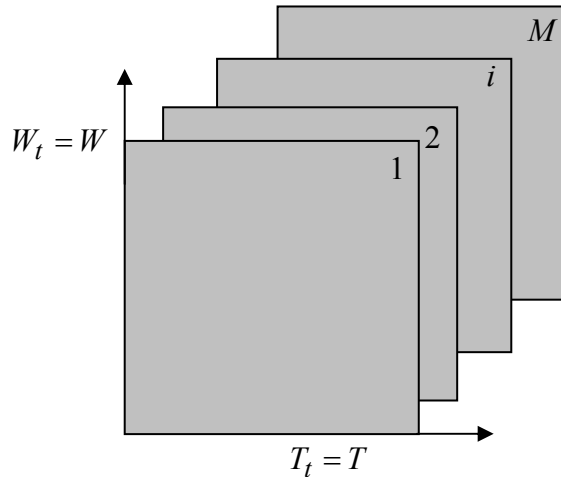


Рис. 2.21. Распределение ресурса при ортогонализации.

В заключении подведем итог материалу, изложенному в 2.3-2.5. Как можно видеть, теоретически классическая задача M -ичной передачи информации не ориентирована на безусловное применение технологии распределенного спектра и, в принципе, оптимальный выбор сигналов может быть осуществлен в классе простых сигналов. С другой стороны, существуют причины реализационного плана вместе с желанием приобрести преимущества, присущие распределенному спектру вне рамок классического приема. Поскольку упомянутые возможности полностью связаны с необходимостью использования сигналов со значительным частотно-временным ресурсом $W_t T_t \gg 1$, то в последнее время разработчики систем склоняются к предпочтению сигналов с распределенным спектром, а не простых.

2.6 Обмен между выигрышем от ортогонального кодирования и шириной полосы.

Оценим теперь выигрыш, который сопровождает применение ортогональных сигналов со случаем отсутствия кодирования, т.е. непосредственной передачей потока информационных бит источника. Предположим, что энергетический ресурс позволяет передавать каждый бит данных с энергией E_b , отсутствует специальное кодирование потока бит данных и каждый информационный бит передается оптимальным образом с помощью пары противоположных сигналов, другими словами используется БФМ. Среди блоков из m последовательных бит возможны любые комбинации, в том числе и такие, которые отличаются друг от друга значением только одного бита. Поэтому минимальная величина квадрата расстояния между сигналами, отвечающими не идентичным m -битовым блокам, совпадает со значением квадрата расстояния между одноканальными противоположными сигналами, т.е. согласно рис. 2.4, а $d_{\min, u}^2 = 4E_b$, где второй индекс обозначает отсутствие кодирования (uncoded).

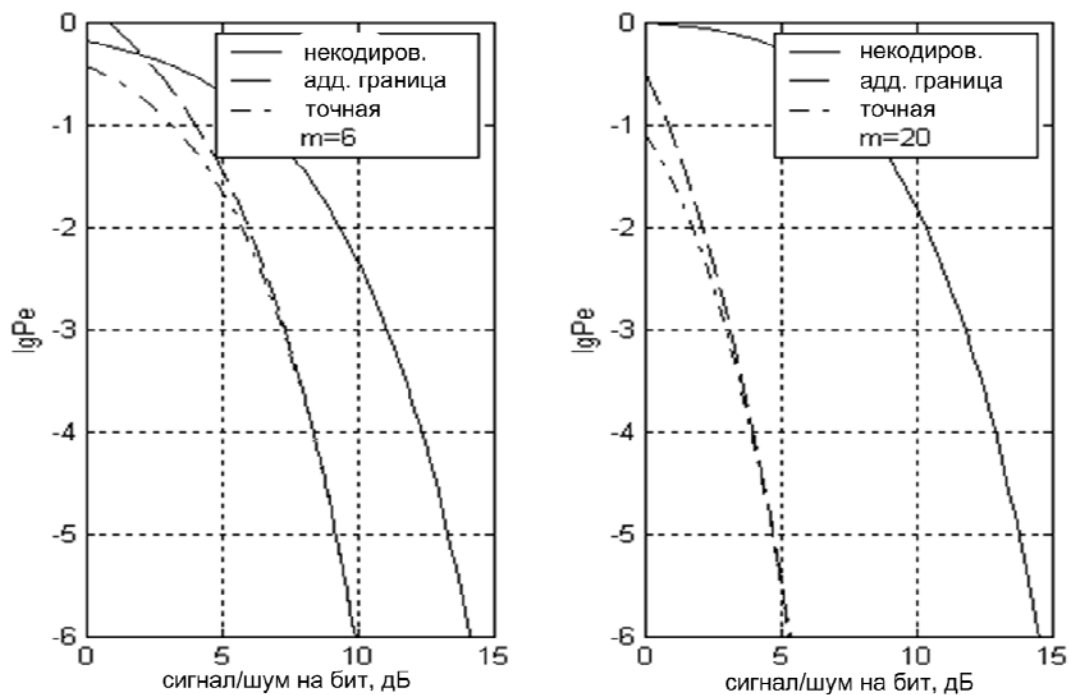


Рис. 2.21. Сравнение вероятности ошибки для некодированной передачи и ортогональной сигнализации

Рассмотрим теперь другую систему, в которой все различные m – битовые блоки передаются ортогональными сигналами. Ясно, что каждый такой сигнал характеризуется энергией, равной mE_b , сохраняя неизменной ее значение на бит E_b . Тогда квадрат расстояния (одинаковый для любой пары сигналов, поскольку ортогональные сигналы эквидистантны) между сигналами снова легко может быть определен на основании рис.2.4, b : $d_{\min, \text{ort}}^2 = 2mE_b$. Очевидно, что выигрыш G_a ортогональных сигналов в величине минимума квадрата расстояния у варианта отсутствия кодирования составляет $m/2$. При альтернативном подходе для обеспечения одинакового минимума расстояния передача данных без кодирования требует энергетических затрат в $G_a = m/2$ больше, чем в случае ортогонального кодирования. Принимая во внимание соотношение (2.29), минимум расстояния асимптотически определяет вероятность ошибки для любых сигналов, используемых в M – ичной передаче. Следовательно, требование высокой достоверности приема, автоматически влекущее за собой необходимость высокого значения отношения сигнал-шум, означает, что одинаковая надежность двух рассмотренных систем возможна только в случае G_a раз больших энергетических затратах для варианта без кодирования. Таким образом, *асимптотический* выигрыш G_a от кодирования ортогональными сигналами является адекватным показателем преимущества ортогональной сигнализации в пределе, т.е. при отношении сигнал-шум, стремящемся к бесконечности.

Для определения величины выигрыша от ортогонального кодирования при конечном значении отношения сигнал-шум, также как скорости его сходимости к асимптотическому случаю на рис.2.21 представлены семейства кривых для двух значений длины m – битового блока: $m=6$ и $m=20$. Первая кривая (сплошная линия) демонстрирует вероятность ошибочного приема некодированного блока, а две других вычислены в предположении кодирования m – битовых блоков $M = 2^m$ ортогональными сигналами, которые обрабатываются когерентно. Пунктирная кривая построена на основании аддитивной границы (2.37), тогда как кривая, обозначенная через точку и тире, отвечает точной формуле вычисления вероятности ошибки когерентного приема M ортогональных сигналов, вывод которой можно найти во многих популярных книгах по основам теории связи [например 1,3]:

$$P_{e,ort} = 1 - \int_{-\infty}^{\infty} \exp\left[-\frac{(x - q_b \sqrt{m})^2}{2}\right] \Phi^{M-1}(x) dx,$$

где $q_b = \sqrt{2E_b / N_0}$ – отношение сигнал-шум на бит, а $\Phi(x) = 1 - Q(x)$ – функция ошибок.

Основной вывод, который следует из графиков на рис. 2.21, состоит в высокой достоверности аддитивной границы. В области типичных для практики требований к вероятности ошибки просвет между величиной отношения сигнал-шум на бит, оцененного из границы и вычисленного точно, составляет менее 0.5 дБ ($P_e \approx 10^{-2}$) и быстро уменьшается с ростом требований к достоверности передачи (менее 0.2 дБ при $P_e \leq 10^{-3}$). В диапазоне значений P_e от 10^{-2} до 10^{-6} действительный выигрыш G от ортогонального кодирования возрастает с 3,5 до 4,2 дБ ($m = 6$) и от 8,5 до 8,9 дБ ($m = 20$). Сравнивая эти цифры с асимптотическими значениями (4.8 и 10 дБ соответственно), нетрудно увидеть более чем хорошее совпадение между ними, оправдывающее использование асимптотического выигрыша от кодирования в качестве первого приближения оценки эффективности ортогонального кодирования.

Очень оптимистическое мнение об ортогональной сигнализации, которое может возникнуть свете приведенных выше результатов, значительно снижается реальной стоимостью выигрыша от кодирования. Последний приобретается за счет расширения полосы, поскольку, как это было установлено в 2.3, размерность сигнального пространства n_s , т.е. число ортогональных сигналов $M = n_s$, непосредственно определяется общим частотно-временным ресурсом $W_t T_t$ системы. Интересуясь в основном установлением количественного порядка и игнорируя тривиальный путь удвоения числа когерентных ортогональных сигналов, имеем $M = W_t T_t$, или $W_t = M / T_t$. Пусть необходимая скорость передачи в системе составляет R бит/с, что соответствует передаче $m = R T_t$ бит за выделенный интервал времени T_t . Ясно, что ортогональное кодирование битовых блоков такой длины может быть осуществлено с помощью $M = 2^m = 2^{R T_t}$ сигналов, обеспечивая асимптотический выигрыш от кодирования $G_a = m / 2 = R T_t / 2$. Тогда спектральная эффективность R / W_t системы, т.е. скорость, приходящаяся на 1 Гц полосы, использующей ортогональные сигналы, будет

$$\frac{R}{W_t} = \frac{R T_t}{2^{R T_t}} = \frac{2 G_a}{2^{2 G_a}}, \quad (2.66)$$

откуда следует ее значительное уменьшение (почти экспоненциальное) с ростом желаемого выигрыша от кодирования.

Обратимся к характеристикам численного примера.

Пример. Для цифровых систем передачи речи очень типичной является скорость $R = 9.6$ кбит/с (мобильная телефония, системы мультимедиа и др.). Предположим, что необходимо уменьшить в три раза (на 4.8 дБ) излучаемую мощность без потери в величине скорости. Если достижение поставленной цели осуществляется путем использования ортогональных сигналов, то это становится возможным лишь за счет снижения спектральной эффективности с 1 до 6/64, как следует из соотношения (2.66). Другими словами, для сохранения неизменной скорости $R = 9.6$ кбит/с необходимо использовать полосу не менее чем 100 кГц. Эта цифра не является запредельной для многих приложений, так, например, на данном принципе строится обратный канал сотовой системы мобильной связи стандарта cdmaOne.

Представим теперь ситуацию, когда, основываясь на результатах успешно решенной предыдущей задаче, планируется достичь десятикратного (10 дБ) уменьшения излучаемой мощности. Реализация поставленной цели требует применения $M = 2^{20} > 10^6$ ортогональных сигналов для кодирования блоков длиной $m = 20$ бит. Это приведет к спектральной эф-

эффективности, меньшей $2 \cdot 10^{-5}$, или полосе, шире 480 МГц, что представляется совершенно неприемлемым при скорости в 9.6 кбит/с.

Вышеприведенное обсуждение свидетельствует об очень жестком характере обмена между энергетической и спектральной эффективностью, присущем ортогональному кодированию. В то же время уместно отметить, что, несмотря на значительный энергетический выигрыш, практически недостижимый при использовании ортогональных сигналов из-за неприемлемых требований к полосе, асимптотический выигрыш при ортогональном кодировании может служить хорошей отправной точкой, будучи верхней границей теоретической эффективности при любом методе кодирования m – битовых блоков.

3 Каналы со случайными параметрами

Большинство реальных систем связи используют каналы передачи, параметры которых изменяются во времени случайным образом. Это могут быть *медленные* изменения, период которых сравним с длительностью сеанса связи или превосходить его. Причинами медленных замираний являются изменения свойств среды распространения радиоволн в зависимости от метеоусловий, времени суток, года, от солнечной активности и т.п. Соответственно, *быстрые* изменения характеризуются периодом, меньшим длительности сеанса связи. В результате этих изменений мощность принимаемых сигналов также меняется – возникают *замирания* (фединг). Если устранить медленные замирания, проявляющиеся от сеанса к сеансу, можно благодаря *системному запасу*, в данном случае энергетическому, то быстрые замирания существенно ухудшают качество связи. Медленные замирания обычно достаточно хорошо описываются логарифмически нормальным распределением.

Наиболее яркими представителями каналов со случайными параметрами являются каналы городской мобильной связи. Электромагнитные волны в таких каналах распространяются по многим путям, причем прямой луч, распространяющийся по *линии визирования* (ЛВ), зачастую отсутствует. Каналы с различными путями распространения электромагнитных волн называют *многолучевыми*. Вследствие интерференции волн различных лучей результирующая сумма в точке приема существенно меняется как при движении приемника относительно передатчика, так и при отсутствии этого движения – из-за отражений от движущихся объектов (транспортных средств, листва в ветреную погоду и т.п.).

Быстрое (в литературе встречаются названия мелкомасштабное, или вызванное изменением микроструктуры среды) замирание — это значительные изменения амплитуды и фазы сигнала, которые на практике могут быть результатом небольших изменений (порядка половины длины волны) расстояния между передатчиком и приемником.

Прием сигналов в каналах со случайными параметрами представляет собой значительно более сложную задачу, чем прием в каналах с постоянными параметрами: замирания на некоторых интервалах приводят к резкому падению уровня принимаемого сигнала, зачастую ниже уровня аддитивных помех. В результате достоверность принятой информации резко ухудшается, а скорость передачи падает.

Ограничимся в дальнейшем рассмотрением многолучевых каналов, характерных для ионосферных, тропосферных, мобильных систем связи, а также гидроакустических систем. Существует два класса таких каналов:

- каналы с прямой волной;
- каналы с отраженно-рассеянной волной.

Каналы, содержащие прямую волну, характеризуются случайными флуктуациями огибающей принятого сигнала вокруг некоторого среднего значения. Плотность распределения огибающей обычно достаточно хорошо описывается законом Релея-Райса.

Каналы, в которых прямая волна отсутствует, имеют более глубокие замирания. Если огибающая принятого сигнала описывается законом Релея, такие каналы называются *релеевскими*, характеризующимися наиболее глубокими замираниями.

Существует три основных механизма, воздействующих на распространение сигнала в системах связи с многолучевым распространением радиоволн.

- *Отражение* происходит тогда, когда распространяющаяся электромагнитная волна сталкивается с гладкой поверхностью, размер которой гораздо больше длины волны радиочастотного сигнала λ .
- *Дифракция* встречается тогда, когда путь распространения между передатчиком и приемником преграждается плотным телом, размеры которого велики по сравнению с λ , что вызывает появление вторичных волн, образующихся позади преграждающего тела. Дифракция — это явление, которое является причиной того, что распространение радиочастотной энергии от передатчика к приемнику происходит в обход пути прямой видимости

между ними. Ее часто называют *затенением*, поскольку дифрагированное поле может достичь приемника, даже если оно затенено непроницаемой преградой.

- *Рассеяние* встречается тогда, когда радиоволна сталкивается с любой неровной поверхностью или с поверхностью, размеры которой порядка λ или меньше, что приводит к распространению (рассеянию) или отражению энергии во всех направлениях. В городской местности обычные препятствия, вызывающие рассеивание сигнала, — это фонарные столбы, уличные знаки и листья. Название *рассеивающий элемент* применимо к любым препятствиям на пути распространения, которые являются причиной отражения или рассеяния сигнала.

Основная терминология в каналах с замираниями

- *Межсимвольная интерференция* (МСИ) – наложение предыдущего символа на последующий.
- *Доплеровское растяжение* спектра сигнала Δf_D вызывается отражением (или прохождением) электромагнитной волны от неоднородностей среды распространения или других объектов, движущихся со случайными скоростями и в случайных направлениях.
- *Полоса когерентности* – полоса частот, в которой доплеровское смещение (не путать с растяжением!) одинаково. Ее также называют полосой частот *гладких замираний*.
- *Селективные замирания* характеризуются случайными доплеровскими смещениями на разных частотах (между замираниями на разных частотах отсутствует корреляция).
- *Общие замирания* – одинаковые изменения на всех частотах.
- *Глубина замираний* описывается изменением уровня огибающей сигнала относительно ее медианного значения.
- *Скорость замираний* – среднее число односторонних пересечений огибающей ее медианного значения в единицу времени.

На рис. 3.1 представлены типичные случаи передаточной частотной функции канала в сравнении со спектром передаваемого сигнала.

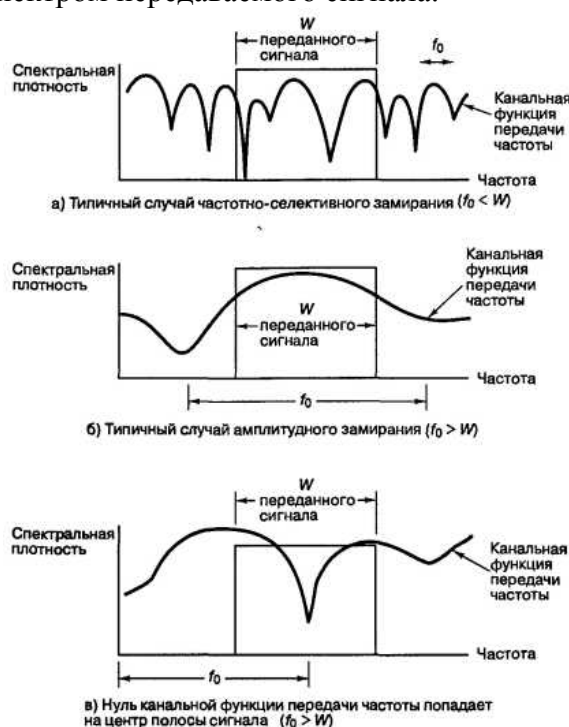


Рис. 3.1. Коэффициент передачи канала и спектр сигнала

Здесь приведено три примера. В каждом из них показана зависимость спектральной плотности от частоты переданного сигнала, имеющего полосу W Гц. На графике (рис. 3.1, а) на сигнал наложена частотная передаточная функция частотно-селективного канала ($f_0 < W$), f_0 – полоса когерентности. На рис. 3.1, а показано, что различные спектральные компоненты переданного сигнала будут подвергаться различному воздействию.

Частотно-неселективное, или амплитудное, ухудшение характеристик происходит тогда, когда $f_0 > W$. Следовательно, все спектральные компоненты сигнала будут подвергаться одинаковому воздействию со стороны канала (например, замирать или не замирать). Это показано на рис. 3.1, б, где изображена спектральная плотность того же переданного сигнала, имеющего полосу W Гц.

Однако на этот сигнал теперь наложена частотная передаточная функция канала с амплитудным замиранием ($f_0 > W$). Из рис. 3.1, а видно, что воздействие на все спектральные компоненты будет приблизительно равным. Амплитудное замирание не приносит искажений, связанных с внесенной каналом МСИ, однако все же стоит ожидать ухудшения характеристик сигнала, выражающегося в уменьшении отношения сигнал/шум (ОСШ). Чтобы избежать искажения вследствие внесенной каналом МСИ, необходимо, чтобы канал проявлял только амплитудное замирание. Это происходит при следующем условии:

$$f_0 > W = 1/T$$

Следовательно, полоса когерентности f_0 устанавливает верхний предел скорости передачи, которую можно использовать, не включая в приемник эквалайзер.

На рис. 3.1, б показано обычное графическое представление амплитудного замирания, когда $f_0 > W$. Однако если мобильный радиоприемник изменит свое местонахождение, некоторое время получаемый сигнал будет подвергаться частотно-селективному искажению, несмотря на то, что $f_0 > W$. Соответствующая иллюстрация приведена на рис. 3.1, в, где нуль частотной передаточной функции канала находится около середины полосы спектральной плотности переданного сигнала. Когда это происходит, видеосигнал может искажаться собственными смещенными низкочастотными компонентами. Одним из последствий этого является отсутствие надежного максимума импульса, составляющего основу синхронизации или предназначенного для выборки фазы несущей, переносимой импульсом. Таким образом, хотя канал (на основе среднеквадратических соотношений) отнесен к каналам с амплитудным замиранием, он может периодически проявлять и частотно-селективное замирание. Стоит отметить, что канал мобильной радиосвязи, классифицированный как канал с амплитудным замиранием, не может все время проявлять амплитудное замирание. Когда f_0 становится намного больше W , все меньший интервал времени реализуется состояние, показанное на рис. 3.1, в. Очевидно, что замирание на рис. 3.1, а не зависит от места в полосе частот сигнала, так что частотно-селективное замирание происходит не эпизодически, а все время.

Для характеристики скорости замираний используют также корреляционную функцию замираний. Нормированную корреляционную функцию замираний $\rho(\tau)$ аппроксимируют функциями

$$\begin{aligned} 1) \quad \rho(\tau) &= \exp\left(-\frac{|\tau|}{\tau_{\delta\ddot{e}}}\right), \\ 2) \quad \rho(\tau) &= \exp\left(-\frac{\tau^2}{2\tau_{\delta\ddot{e}}^2}\right), \end{aligned}$$

где $\tau_{\delta\ddot{e}}$ – параметр, однозначно связанный со скоростью замираний. Время корреляции процесса замираний определим, как

$$\tau_k = \int_0^{\infty} \rho(\tau) d\tau.$$

Тогда для приведенных корреляционных функций будем иметь:

- 1) $\tau_k = \tau_{\phi l}$ и $\rho(\tau_k) \approx 0,4$;
- 2) $\tau_k = \tau_{\phi l} \sqrt{\pi} / 2 \approx 1,25 \tau_{\phi l}$ и $\rho(\tau_k) \approx 0,45$.

Интервал корреляции τ_k можно рассматривать как средний период замираний. Обычно он значительно превышает длительность символа T : $\tau_k \gg T$.

Сигнал многолучевого канала (индекс “х” обозначает принадлежность к случайной величине) можно представить в виде

$$s_x(t) = \sum_k s_{ik}(t) = \sum_k S_{ik}(t) f_i[t - \tau_{zk}(t)] \quad (3.1)$$

где k – номер луча, $k = (1, K)$; K – общее число лучей;

S_{ik} – огибающая i -й посылки, принятой по k -му лучу;

τ_{zk} – запаздывание посылки k -го луча;

f_i – посылка единичной амплитуды, соответствующая выбранному виду манипуляции и передаваемому символу. Например, для БФМ:

$$f_i(t) = \begin{cases} \sin(\omega_0 t + \varphi_i), & 0 \leq t \leq T \\ 0, & \text{в остальных случаях} \end{cases}$$

Принятый сигнал (3.1) имеет две компоненты:

$$s_x(t) = s_p(t) + s_{\phi l}(t), \quad (3.2)$$

$s_p(t)$ – регулярная, $s_{\phi l}(t)$ – случайная.

Мгновенные значения сигнала $s_x(t)$ имеют нормальное распределение. Для огибающей S_X используют законы:

Релея-Райса

$$W(S_X) = \frac{S_X}{\sigma_{\phi l}^2} \exp\left(-\frac{S_X^2}{2\sigma_{\phi l}^2}\right) I_0\left(\frac{S_p S_X}{\sigma_{\phi l}^2}\right), \quad \{S_X^2 = 2\sigma_{\phi l}^2\} \quad (3.3)$$

и Накагами (m -распределение) – более общий закон

$$W(S_X) = \frac{2}{\Gamma(m)} \left(\frac{m}{S_X^2}\right)^m I_0\left(-\frac{m S_X^2}{S_X^2}\right), \quad S_X \geq 0, \quad m \geq 0,5 \quad (3.4)$$

Он позволяет описать более широкий класс распределений, нежели закон Релея-Райса в зависимости от параметра m :

$m = 0,5$ – нормальный закон

$m = 1$ – распределение Релея

$m > 1$ – распределение Релея-Райса

Если ни одна составляющая сигнала не имеет регулярной части, то описание производится законом Релея. Такой канал, как отмечалось выше, называется релеевским и характеризуется наиболее глубокими замираниями.

Рассмотренная упрощенная модель замираний сигнала в ряде случаев хорошо согласуется с многочисленными результатами экспериментальных исследований, а также с данными практических наблюдений. Дальнейшее рассмотрение будем проводить в предположении, что замирания являются общими. Общие замирания не изменяют форму сигнала и поэтому их можно трактовать как результат умножения сигнала на некоторую случайную функцию времени, отображающую процесс замираний в канале. Как отмечалось ранее, эту функцию называют мультипликативной помехой, а соответствующие ей замирания – мультипликативными. Будем полагать, что эти замирания относятся к классу, описываемом обобщенным законом Релея.

3.1 Вероятность ошибки приема флуктуирующих сигналов

Запишем сигнал на входе приемника в виде

$$s_x(t) = \alpha S_0 \sin(\omega_0 t + \varphi_x), \quad 0 \leq t \leq T \quad (3.5)$$

Здесь $\alpha = S_x / S_0$ – случайный мультипликативный коэффициент, S_x – случайная амплитуда, S_0 – амплитуда в отсутствие замираний.

Тогда

$$h_x^2 = \frac{E_x}{N_0} = \frac{S_x^2 T}{2N_0} = \alpha^2 S_0^2 \frac{T}{2N_0}; \quad h_x = \alpha S_0 \sqrt{\frac{T}{2N_0}} \quad (3.6)$$

Величина h_x характеризует ОСШ принятого в данный момент времени символа. Вероятность ошибки при этом обозначим как $P_E(h_x)$. Необходимо найти среднюю за сеанс связи вероятность ошибки. Математическое ожидание $M[P_E(h_x)]$ ищем обычным образом:

$$\overline{P_E} = \int_0^\infty P_E(h_x) W(h_x) dh_x, \quad (3.7)$$

где $W(h_x)$ – неизвестное пока распределение случайной величины h_x . Будем предполагать, что канал является релеевским, т.е. амплитуда S_x распределена по релеевскому закону. Связь между величинами h_x и S_x дает соотношение (3.6). Представим его в виде

$$h_x^2 = \frac{S_x^2 T W}{2N_0 W} = \frac{S_x^2 B_c}{2\sigma_\theta^2} = \frac{P_{cx}}{P_\theta} B_c, \text{ или} \\ h_x = \frac{S_x \sqrt{B_c}}{\sigma_{ui} \sqrt{2}} = a S_x, \quad (3.8)$$

где $B_c = WT$ – база сигнала, σ_θ^2 – мощность шума, P_{cx} – мощность принятого сигнала.

Распределение случайной амплитуды запишем в виде:

$$W(S_x) = \frac{S_x}{\sigma_{\delta\theta}^2} \exp\left(-\frac{S_x^2}{2\sigma_{\delta\theta}^2}\right), \quad (3.9)$$

где $\overline{P_{\phi l}} = 0,5 \overline{S_{\phi l}^2} = 2\sigma_{\phi l}^2$ – мощность флуктуирующей составляющей.

Воспользуемся теоремой о функциональном преобразовании распределений случайных величин, которая в данном контексте выглядит, как

$W(y) = W_x[\varphi(y)] [\varphi'(y)]$, где:

$y = f(x)$ – связь между старой и новой случайными величинами x и y : $h_x = a S_x$,

$\varphi(y)$ – функция, обратная $f(x)$: $S_x = h_x / a$,

$W_x(x)$ – плотность распределения вероятностей случайной величины x . Тогда:

$$W(h_x) = \frac{h_x}{a \sigma_{\phi l}^2} \exp\left(-\frac{h_x^2}{2a^2 \sigma_{\phi l}^2}\right) \frac{1}{a} = \frac{2h_x}{h_x^2} \exp\left(-\frac{h_x^2}{h^2}\right) \quad (3.10)$$

$$\overline{h_x^2} = 2a^2 \sigma_{\phi l}^2 = 2 \frac{B_c \sigma_{\phi l}^2}{\sigma_{ui}^2 \cdot 2} = B_c \frac{\sigma_{\phi l}^2}{\sigma_{ui}^2} = h^2$$

В последнем соотношении использовано равенство средней энергии посылки флуктуирующего сигнала и энергии посылки сигнала в отсутствии флуктуаций, что следует из физических соображений. Говоря другими словами, $\overline{\alpha^2} = 1$.

Окончательно, искомую плотность распределения запишем в виде

$$W(h_x) = \frac{2h_x}{h^2} \exp\left(-\frac{h_x^2}{h^2}\right) \quad (3.11)$$

Теперь имеется все необходимое для вычисления вероятности ошибки в каналах со случайными параметрами. Для получения удобной сокращенной записи введем коэффициент γ_c , учитывающий тип используемых сигналов:

$\gamma_c = \sqrt{2}$ для противоположных сигналов,

$\gamma_c = 1$ для ортогональных сигналов,

$\gamma_c = 1/\sqrt{2}$ для сигналов БАМ

3.1.1 Когерентный прием двоичных сигналов

Среднее значение вероятности ошибки в соответствии с (3.7) ищем в виде

$$\overline{P_E} = \int_0^\infty Q(\gamma_c h_x) \frac{2h_x}{h^2} \exp\left(-\frac{h_x^2}{h^2}\right) dh_x \quad (3.12)$$

Интеграл (3.12) берем по частям, воспользовавшись правилом $\int u dv = uv - \int v du$.

В нашем случае

$u = Q(\gamma_c h_x)$, $dv = \exp(-h_x^2/h^2) d(h_x^2/h^2)$;

$du = \frac{dQ(\gamma_c h_x)}{dh_x} = \frac{\gamma_c}{\sqrt{2\pi}} \exp\left(-\frac{\gamma_c^2 h_x^2}{2}\right)$, $v = \exp(-h_x^2/h^2)$.

$$\begin{aligned} \overline{P_E} &= Q(\gamma_c h_x) \exp(-h_x^2/h^2) \Big|_0^\infty - \frac{1}{\sqrt{2\pi}} \int_0^\infty \gamma_c \exp\left(-\frac{\gamma_c^2 h_x^2}{2}\right) \exp\left(-\frac{h_x^2}{h^2}\right) dh_x = \\ &= 0,5 \left(1 - \frac{\gamma_c h}{\sqrt{2 + \gamma_c^2 h^2}} \right) \end{aligned} \quad (3.13)$$

При вычислении (3.13) использован табличный интеграл [6]

$$\int_0^\infty e^{-r^2 x^2} dx = \frac{\sqrt{\pi}}{2r}$$

Цифровые системы связи должны работать с вероятностями ошибок $P_E \leq 10^{-3}$, что обеспечивается отношением сигнал/шум $h \geq 3$. В этом случае

$$\overline{P_E} \approx 1 / 2 \gamma_c^2 h^2 \quad (3.14)$$

и для различных типов манипуляции получаем:

$\overline{P_E} \approx 1 / h^2$, $\gamma_c = 1 / \sqrt{2}$ для сигналов БАМ;

$\overline{P_E} \approx 1 / 2 h^2$, $\gamma_c = 1$ для сигналов БЧМ;

$\overline{P_E} \approx 1 / 4 h^2$, $\gamma_c = \sqrt{2}$ для сигналов БФМ.

3.1.2 Некогерентный прием двоичных сигналов

Ранее были получены выражения для вероятности ошибки некогерентного приема сигналов БАМ и БЧМ (2.56) и (2.58). Объединяя их с помощью коэффициента γ_c , получим заменой h на h_x :

$$P_E(h_x) = 0,5 \exp(-\gamma_c^2 h_x^2 / 2) \quad (3.15)$$

Проделав вычисления, аналогичные разделу 3.1.1, получим выражения для средней вероятности ошибки в канале со случайными параметрами:

$$\overline{P_E} = 1 / (2 + \gamma_c^2 h^2), \quad (3.16)$$

которое упрощается для больших отношений сигнал/шум до

$$\overline{P_E} \approx 1 / \gamma_c^2 h^2. \quad (3.17)$$

Для различных типов манипуляции будем иметь:

$$\overline{P_E} \approx 2/h^2, \quad \gamma_c = 1/\sqrt{2} \quad \text{для сигналов БАМ};$$

$$\overline{P_E} \approx 1/h^2, \quad \gamma_c = 1 \quad \text{для сигналов БЧМ};$$

$$\overline{P_E} \approx 1/2h^2, \quad \gamma_c = \sqrt{2} \quad \text{для сигналов БФМ}.$$

Весьма поучительно сравнение затрат энергии для получения вероятности ошибок, не превышающих 10^{-5} , при отсутствии и наличии замираний, которое представлено таблицей 3.1.

Таблица 3.1

Сигнал	h^2	
	Гауссовский	Релеевский
БАМ	45	$2 \cdot 10^5$
БЧМ	22	10^5
БФМ	11	$5 \cdot 10^4$

Из приведенных в таблице расчетных данных следует, что без принятия специальных мер некогерентный прием при наличии замираний требует огромных энергетических затрат, совершенно неприемлемых в практической реализации. Поэтому рассмотрим далее методы, применяемые для борьбы с замираниями.

3.2 Разнесение сигналов

Разнесение сигналов является эффективным методом повышения качества передачи информации в каналах со случайными параметрами. Суть метода состоит в образовании в точке приема нескольких некоррелированных замирающих копий сигнала. Тогда, если вероятность того, что амплитуда i -й копии меньше некоторой величины U_0 , равна $p(S_{Xi} < U_0) = p_i$, то вероятность того, что амплитуды всех образцов сигнала также будут меньше U_0 , будет равна

$$p(S_{X1} < U_0, S_{X2} < U_0, \dots, S_{Xn} < U_0) = \prod_{i=1}^n p_i. \quad (3.18)$$

Если статистика замираний всех копий сигнала одинакова $p_i = p, i = (1, n)$, то вероятность события (3.18) становится равной

$$p_{\text{замир}} = \prod_{i=1}^n p_i = p^n. \quad (3.19)$$

Из выражений (3.18), (3.19) следует, что с увеличением числа копий n вероятность одновременных сильных замираний резко уменьшается, что и используется для ослабления влияния замираний.

Возможны различные варианты организации независимых (некоррелированных) ветвей разнесения. Их можно классифицировать следующим образом:

- *пространственное* разнесение;
- *частотное* разнесение;
- *временное* разнесение;
- *угловое* разнесение;
- *поляризационное* разнесение.

Пространственное разнесение проще всего получить за счет нескольких приемных антенн, расположенных на расстоянии 7 – 10 длин волн друг от друга. При этом замирания в отдельных антеннах можно считать независимыми. Другой положительной стороной является энергетический выигрыш, получаемый при объединении сигналов отдельных антенн. Применение нескольких антенн в пункте передачи не имеет столь явных преимуществ, так

как в приемник будут поступать многолучевые сигналы от каждой передающей антенны. Их обработка возможна, но ее описание выходит за рамки настоящего учебного пособия.

Частотное разнесение предполагает передачу сигнала одновременно на разных частотах. Чтобы обеспечить независимость ветвей разнесения, расстояния между частотами должно превышать полосу когерентности. Отсюда следует основной недостаток этого вида разнесения: его реализация требует существенных частотных затрат.

Временное разнесение обеспечивается повторением передачи сообщения, причем число копий сообщения ограничивается требуемой скоростью передачи, которая падает во столько раз, сколько используется копий сигнала.

Угловое разнесение использует тот факт, что отдельные лучи приходят в точку приема по разным направлениям. Для разделения лучей необходимо применять остронаправленные антенны и знать направления прихода. Поэтому угловое разнесение практически не используется.

Поляризационное разнесение основано на слабой корреляции волн с ортогональной поляризацией (вертикальной и горизонтальной) и пока не нашло широкого применения.

3.2.1 Широкополосные методы борьбы с многолучевостью

Полную информацию о наличии многолучевого распространения радиоволн и его параметрах (количество лучей K , интенсивность a_k , запаздывание t_k , фазовый сдвиг огибающей θ_k), дает импульсная характеристика канала связи h_c . Для ее получения необходимо, чтобы передатчик излучал настолько короткий одиночный импульс, что на приемной стороне все многолучевые сигналы были бы разрешимы. На практике, однако, достаточно иметь длительность импульса передатчика T_p не более обратной величины полосы частот, занимаемой системой связи: $T_p \leq 1/W$.

На рис. 3.2 представлена осциллограмма принятого многолучевого сигнала в деловом квартале Сан-Франциско. Ее обычно называют *профилем многолучевости*. Передатчик излучал радиоимпульсы длительностью 0,1 мкс.



Рис. 3.2. Профиль многолучевости делового квартала Сан-Франциско [8]

Видно, что при этом не все лучи разрешаются; часть из них, отраженная от протяженных объектов, сливается в импульсы длительностью более 0,1 мкс.

Запаздывания сигналов отдельных лучей измеряются относительно сигнала, приходящего по линии визирования (ЛВ): передатчик – приемник (прямой луч). Его запаздывание обозначают как ЛВЗ –

запаздывание вдоль линии визирования. Профиль многолучевости есть не что иное, как действительная часть импульсной характеристики канала $\text{Re}\{h_c\}$.

Применительно к широкополосным системам, которым будет уделено основное внимание, целесообразно положить, что все лучи разрешимы, т. е. выполняется условие

$$|t_k - t_l| > 1/W \quad \text{для всех } k \neq l. \quad (3.20)$$

Любые два луча, например, k_1 и k_2 , для которых $|t_{k1} - t_{k2}| < 1/W$, будут рассматриваться как один луч с общим запаздыванием $t_k \approx t_{k1} \approx t_{k2}$, интенсивность и фазовый сдвиг которого определяются выражением

$$a_k \exp(j\theta_k) \approx a_{k1} \exp(j\theta_{k1}) + a_{k2} \exp(j\theta_{k2}).$$

Прием сигналов в условиях многолучевости распространения представляет собой одну из форм разнесенного приема, где информация от передатчика к приемнику передается по нескольким естественно разнесенным лучам, а не по нескольким заранее разнесенным частотным, временным, пространственным и т. п. каналам. Следовательно,

вместо того, чтобы рассматривать многолучевость распространения как некое вредное явление, влияние которого необходимо так или иначе устранить, его можно рассматривать как некое благоприятное явление, которое можно использовать для улучшения характеристик системы.

Опубликованные в литературе работы по расчету приемников многолучевых сигналов можно разбить на две основные группы. В наиболее ранних из них (см. библиографию в обзоре [8]) главное внимание уделялось структуре приемника при наличии явно разрешимых лучей; использование такой схемы приема обосновывалось возможностью оптимального объединения вкладов отдельных лучей. В своей простейшей форме этот подход пренебрегает наличием межсимвольной интерференции, которая может появиться в том случае, когда из-за запаздываний сигнала, обусловленных многолучевостью распространения, отклик приемника на переданный символ будет попадать на временные интервалы, занятые последующими переданными символами. Поэтому такой подход оправдан лишь в том случае, когда длительность переданного символа велика по сравнению с длительностью профиля многолучевости (длительностью многолучевого сигнала).

В более поздних работах было показано, что методы, разработанные для коррекции данных, передаваемых по телефонным линиям, применимы и для решения проблемы многолучевости в системах радиосвязи. Здесь при конструировании приемника основное внимание уделяется уменьшению эффектов межсимвольной интерференции, а объединение сигналов, пришедших по различным лучам, подразумевается лишь в неявной форме. Такой подход, по-видимому, наиболее целесообразен в том случае, когда отдельные лучи распространения сигнала не разрешаются, а длительность символа много меньше времени многолучевого растяжения сигнала.

Под временем многолучевого растяжения сигнала τ_p , которое выше определялось как длительность профиля многолучевости, будем понимать разницу между максимальным и минимальным запаздыванием лучей:

$$\tau_p = \max \tau_{зан} - \min \tau_{зан}$$

3.2.1.1 Оптимальный приемник при многолучевом распространении сигнала

Начнем с простейшего случая, когда канал связи образован только одним лучом, т.е. положим в (3.1) $K = 1$. Предположим сначала, что интенсивность луча a_k и его запаздывание t_k известны (для упрощения положим $t_0 = 0$), а фаза несущей θ_0 не известна и является случайной переменной, равномерно распределенной на интервале $[0, 2\pi]$.

Поскольку отсутствие многолучевости подразумевает отсутствие межсимвольной интерференции, то основное внимание мы уделим приему отдельного символа, скажем, на интервале $0 \leq t < T$. Знание этого интервала подразумевает, конечно, наличие некоторой процедуры синхронизации приемника.

Оптимальный приемник, т.е. приемник, который с минимальной вероятностью ошибки принимает решение о том, какой из сигналов s_i , $i = (1, M)$ передан, имеет, как известно, структуру, показанную на рис. 3.3. Это некогерентный приемник: фаза любого луча полагается случайной, так как при приеме сигналов подвижной станцией она изменяется слишком быстро, для того чтобы можно было применить когерентные методы приема.

Из рисунка видно, что принятый сигнал $y(t)$ пропускается через набор из M фильтров, каждый из которых согласован с одним из M сигналов, т.е. через фильтры с импульсными характеристиками $s_i(T-t)$, $0 \leq t < T$. К выходам фильтров подключены детекторы огибающей, выходные напряжения которых сравниваются в момент времени $t = T$. Индекс $i=1, \dots, M$ наибольшего из них является выходом приемника.

Изображенный на рис. 3.3 случай соответствует переданному сигналу с индексом “ j ”, шумовая составляющая принятого сигнала полагается пренебрежимо малой.

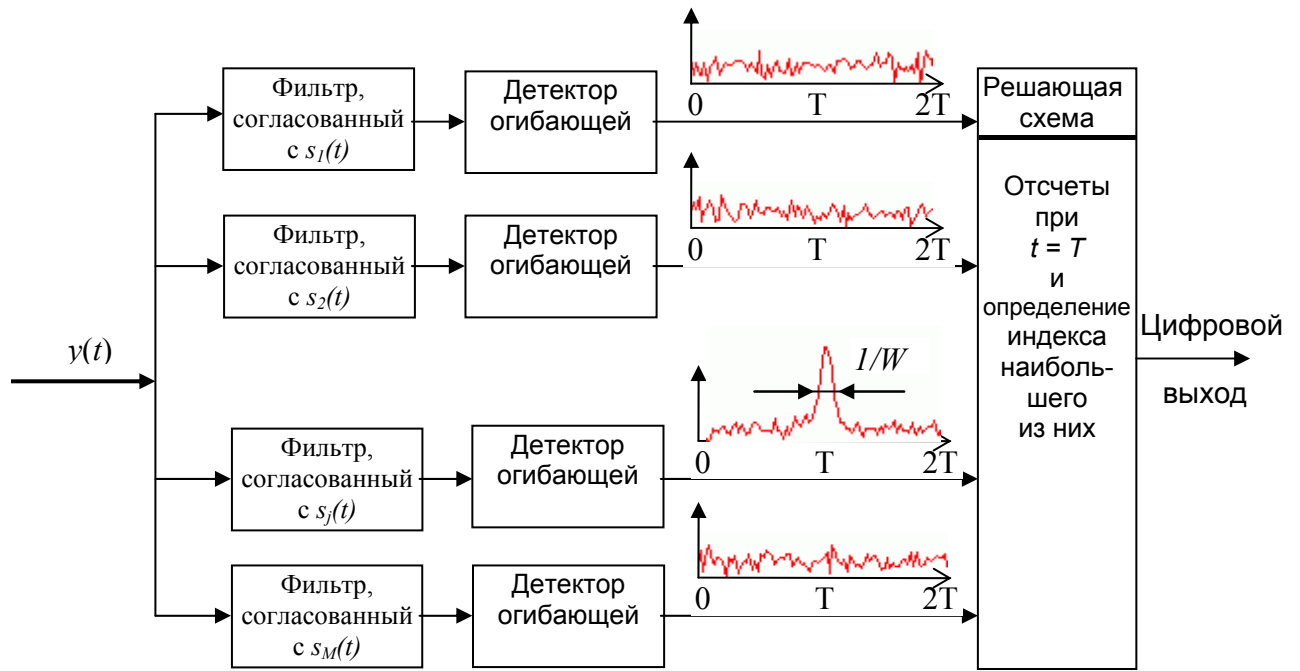


Рис. 3.3. Оптимальный некогерентный приемник для однолучевого канала (равновероятные сигналы равной энергии; гауссовский шум).

Главный пик на выходе детектора j -го канала имеет амплитуду, равную энергии сигнала E в отсутствии шумов; сам сигнал полагается широкополосным (сложным) с базой $B_c = WT \gg 1$, так что согласованный фильтр сжимает его во времени до длительности $1/W$. В окрестности $|t| > 1/W$ j -го канала, а также в интервале $0 \leq t < 2T$ остальных каналов сигнал представляет собой взаимокорреляционные функции принятого сигнала с импульсными характеристиками согласованных фильтров

$$R_{ik} \propto \int_0^T S_i^*(\tau) S_k(\tau - t) d\tau, \quad i, k = (1, M), \quad |t| \leq T, \quad (3.21)$$

показанные на рисунке в виде шумовых дорожек. Напомним, что S_i – огибающая i -го сигнала. Для реальных сигналов должны выполняться

$$\begin{aligned} |R_{ik}| &\ll E, & k \neq i; \\ |R_{ii}| &= E, & |t| < 1/W. \end{aligned} \quad (3.22)$$

Эти условия гарантируют, что огибающая j -го канала будет иметь острый главный пик, окруженный низкими «боковыми лепестками», а остальные выходные сигналы будут содержать только боковые лепестки низкого уровня. При тщательном выборе сигналов выполнение условий (3.22) означает, что максимальный уровень боковых лепестков у всех сигналов будет примерно в $2\sqrt{WT}$ раз ниже уровня главного лепестка.

Наконец, отметим еще одну важную особенность рис. 3.3. Так, выходные сигналы на нем изображены для одного отдельного символа, переданного в течение временного интервала $0 \leq t < T$. Если другой символ, скажем, $s_i(t)$, был передан сразу после него на временном интервале $T \leq t < 2T$, то, очевидно, отклик на него появится на временном интервале $T \leq t < 3T$. Пик главного лепестка i -го выходного сигнала будет точно соответствовать моменту времени $t=2T$, т. е. появится тогда, когда все отклики на первый символ уже затухнут. Следовательно, беря отсчеты выходных сигналов в момент времени $t=2T$, можно вынести решение относительно одного второго символа, что и подтверждает наше утверждение об отсутствии межсимвольной интерференции в однолучевом случае.

3.2.1.2 Оптимальный приемник многолучевого сигнала. Случай известных запаздываний

Если мы попытаемся использовать приемник, показанный на рис. 3.3, в случае, когда имеется несколько лучей ($K > 1$), то вследствие линейности среды и наличия согласованных

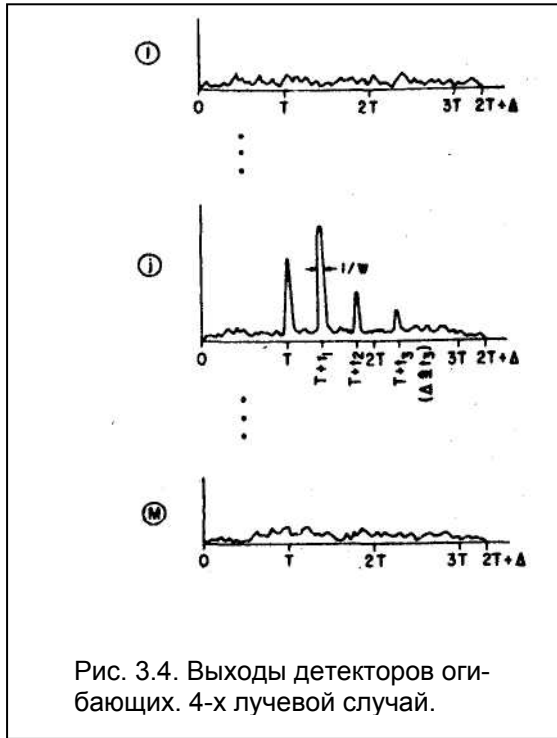


Рис. 3.4. Выходы детекторов огибающих. 4-х лучевой случай.

фильтров можно ожидать, что напряжения на выходах детекторов огибающей будут выглядеть примерно так, как показано на рис. 3.4, на котором они изображены для случая $K = 4$. Здесь через Δ обозначено время многолучевого растяжения: $\Delta = \tau_p$.

Огибающая сигнала в l -м канале, согласованном с пришедшим сигналом, в отсутствии шума равна

$$e_l(t) = \left| \sum_{k=0}^{K-1} a_k \exp(j\theta_k) \int_0^T S_j^*(\tau) S_l(\tau + T + t_k - t) d\tau \right|, \quad (3.23)$$

$$l = (1, M), \quad 0 \leq t < 2T + \Delta.$$

При выполнении условия разрешимости (3.20) главные пики огибающей $e_l(t)$ различны и соответствуют моментам времени $t_0=0$, t_1 , t_2 , и $t_3 = \Delta$. Высоты этих пиков пропорциональны интенсивностям лучей a_k . Боковые лепестки и сигнала $e_j(t)$, и других выходных сигналов (ни один из последних не имеет главного пика) представляют собой смесь боковых лепестков сигналов, пришедших по нескольким лучам. Еще раз напомним, что рис. 3.4 соответствует

передаче только одного сигнала $s_j(t)$, $0 \leq t < T$.

Сигналы, показанные на рис. 3.3 и 3.4, имеют несколько важных различий.

1) Огибающая $e_l(t)$ имеет несколько сильных пиков. Если решающей схеме приемника, показанного на рис. 3.3, известны времена запаздываний $t_0, \dots, t_{K-1}$, то она может объединить отсчеты всех выходных сигналов, тем самым обеспечивая приемнику преимущества разнесенного приема, о котором говорилось выше. Возможность разрешения лучей по сигналам, показанным на рис. 3.4, составляет суть широкополосного подхода. Если бы мы имели $TW \approx 1$, то пики на рис. 3.4 слились бы и идея разнесенного приема (т. е. объединение сигналов) в явном виде была бы больше не применима.

2) Уровни боковых лепестков всех выходных сигналов возрастают, что, согласно (3.23), обусловлено дополнительными вкладами из-за многолучевости.

3) Отклики на символ, переданный на интервале $0 \leq t < T$, теперь делятся более времени $t=2T$ и, следовательно, перекрываются откликами на следующий символ, переданный на интервале $T \leq t < 2T$. Иными словами, из-за наличия многолучевости мы будем иметь дело с межсимвольной интерференцией.

Эффекты 2 и 3 оказывают вредное действие, эффект 1 является благоприятным. Однако, как будет показано ниже, выгоды, обеспечиваемые эффектом 1, значительно превосходят вредные последствия, вызванные эффектами 2 и 3.

Пока же мы будем пренебрегать эффектами межсимвольной интерференции и рассмотрим структуру приемника, для оптимального приема отдельного символа в условиях многолучевости распространения. Предположим сначала, что времена запаздывания сигналов $\{t_k\}_0^{K-1}$ для всех лучей известны. Кроме того, будем снова полагать, что фазы сигналов случайны, независимы и равномерно распределены на интервале $[0, 2\pi]$, а интенсивности лучей $\{a_k\}_0^{K-1}$ — также случайны и, возможно, имеют различные распределения.

Интуитивно ясно, что при наложенных условиях оптимальный приемник все еще будет иметь форму, показанную на рис. 3.3, однако решающая схема должна теперь брать отсчеты

каждого своего входного сигнала в моменты времени $T+t_k$, $k=0, \dots, K-1$, объединять их для каждого входного сигнала и затем сравнивать полученные величины для определения индекса наибольшей из них. Именно так и обстоит дело при выполнении условий (3.20), когда амплитуды импульсов на выходе j -го канала (рис. 3.4) различны. Однако закон оптимального объединения отсчетов несколько усложняется, так как теперь он зависит от статистик лучевых интенсивностей.

Пусть x_{ik} — отсчет i -й огибающей, взятый в момент времени $T+t_k$. (Согласно (3.23), при отсутствии шума $x_{ik} = e_i(t_k)$). Тогда, если интенсивности лучей известны, оптимальный закон объединения отсчетов будет определяться выражением [8]

$$w_i = \sum_{k=0}^{K-1} \ln I_0 \left(\frac{2a_k x_{ik}}{N_0} \right), \quad (3.24)$$

где I_0 — функция Бесселя, а N_0 — плотность мощности шума в канале. Если интенсивность k -го луча имеет релеевское распределение со среднеквадратичным значением $\psi_k = E[a_k]^2$, а интенсивности всех лучей независимы, то оптимальный закон объединения отсчетов будет иметь следующий вид [8]:

$$w_i = \sum_{k=0}^{K-1} \frac{\psi_k x_{ik}^2}{N_0 + \psi_k E}, \quad (3.25)$$

где E — энергия сигналов. Возможны другие, более сложные законы объединения отсчетов для других законов распределения интенсивностей. Однако во всех случаях решение принимается на основе сравнения величин w_i и определения наибольшего из них.

Заметим, что в предложенных законах объединения отсчеты берутся с различными весами. Так, например, в (3.24), отсчеты объединяются приблизительно по линейному закону, так как при больших значениях аргумента функция $I_0(\bullet)$ растет примерно по экспоненциальному закону; однако отсчеты, соответствующие лучам с большой интенсивностью сигнала, берутся с большим весом. В (3.25) отсчеты объединяются по среднеквадратичному закону, т. е. больший вес придается здесь отсчетам, имеющим большие значения, однако отсчеты всех сигналов с большой среднеквадратичной интенсивностью ($\psi_k E \gg N_0$) берутся с равными весами, а отсчеты слабых сигналов подавляются. Таким образом, любой оптимальный закон объединения состоит в относительном подчеркивании более надежных данных и в относительном подавлении менее надежных.

3.2.1.3 Оптимальный приемник многолучевого сигнала. Случай неизвестных запаздываний

Как отмечалось ранее, времена запаздывания лучей $\{t_k\}_0^{K-1}$ и число лучей K часто являются случайными и заранее не известными переменными. Для того чтобы определить структуру оптимального приемника в этом случае, мы несколько упростим модель многолучевости. Теперь мы будем полагать, что времена t_k независимы и описываются одним общим распределением плотности вероятностей $p(t_k)$, $0 < t_k < \Delta$, а их индексы переупорядочены в соответствии с ростом величины t_k . Такая модель противоречит принятым допущениям, потому что условие разрешимости лучей (3.20) в данном случае уже не выполняется, так как возможны два луча, для которых $|t_k - t_l| < 1/W$. Но вероятность этого события будет мала, если мала суммарная вероятность интервалов длительностью, равной или меньшей $1/W$; следовательно, распределение $p(t_k)$ должно иметь размытый («диффузный») характер, не иметь высоких пиков и удовлетворять условию $W\Delta \gg 1$.

Однако для отыскания структуры нашего квазиоптимального приемника указанные отличия от реальности несущественны. Поэтому для случая неизвестных запаздываний интенсивности любых двух лучей a_k и a_l для всех $k \neq l$ будут полагаться независимыми, что отличается от реальности, так как лучи с незначительно различающимися запаздываниями имеют, как правило, коррелированные интенсивности. Однако интуитивно ясно, что, как и в случае с уп-

рощенной моделью запаздываний, приемник, полученный на основе этой упрощенной модели интенсивностей лучей, будет близок к оптимальному.

Такой приемник, с учетом сделанных допущений, должен вычислять величины

$$w_i = \int_T^{T+\Delta} p(t-T) F[X_i(t)] dt, \quad l=(1, M) \quad (3.26)$$

где $p(\bullet)$ — определенная выше плотность распределения времен запаздывания лучей, $X_l(\bullet)$ — огибающая выхода l -го фильтра, а $F[\bullet]$ — некоторая нелинейная функция. Затем значения этих величин сравниваются и индекс наибольшей из них считается принятым сигналом. Таким образом, величину w_i можно реализовать, пропуская $X_i(t)$ сначала через нелинейность F , а затем через фильтр с импульсной характеристикой $h_p(t)=p(\Delta-t)$ и беря отсчет сигнала на выходе этого фильтра в момент времени $t=T+\Delta$. Заметим, что фильтр $h_p(t)$ в идеале должен быть согласован с импульсной характеристикой канала $h_c(t)$: $h_p(t) = h_c(\Delta-t)$.

Решающая схема, показанная на рис. 3.3, в этом случае заменяется схемой, показанной на рис. 3.5.

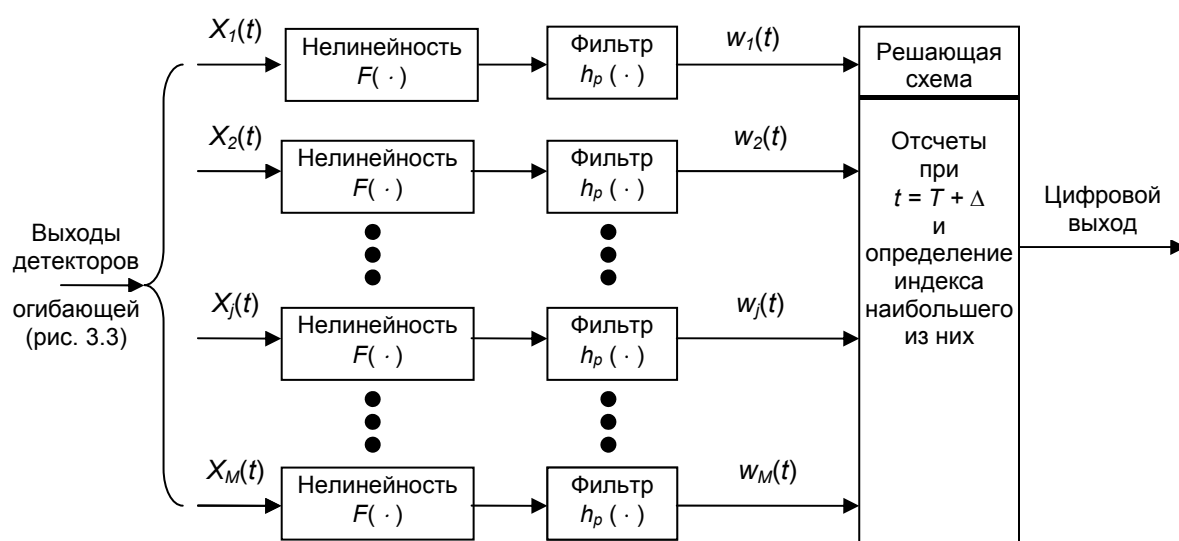


Рис. 3.5. Оптимальный некогерентный приемник для многолучевого канала с неизвестными временами запаздывания лучей.

Следует отметить, что в отличие от выражений (3.24) и (3.25), где отсчеты x_{ik} , $k=0, \dots, K-1$, объединяются по линейному и квадратичному законам, для объединения отсчетов в (3.26) используются экстремальные нелинейности экспоненциального типа, показанная на рис. 3.6.

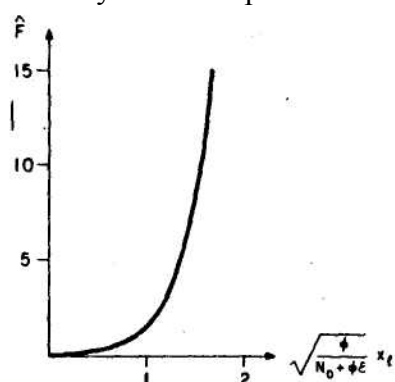


Рис. 3.6. Используемый тип нелинейности $F(\cdot)$

$$F(X_i) = \exp \left[\frac{\psi x_i^2}{N_0 + \psi E} \right] - 1 \quad (3.27)$$

Такая нелинейность подчеркивает сильные сигналы и подавляет слабые. В этом проявляется определенная «самоадаптивность» приемника, в неявном виде присутствующая в (3.26). Следовательно, сигналы на выходе нелинейностей несут в себе информацию, непосредственно характеризующую апостериорные вероятности существования лучей. Согласно (3.26), эти сигналы далее взвешиваются в соответствии с априорными вероятностями размещения лучей, в неявном виде входящими в функцию $p(\bullet)$. Следует подчеркнуть, что изложенная позитивная роль нелинейности проявляется при относительно большой

средней величине ОСШ, когда пики сигнала имеют большую величину по сравнению с шумом и уровнем боковых лепестков. При малом ОСШ эффект подавления шума нелинейностями будет выражаться слабее — тем слабее, чем меньше ОСШ. Можно показать, что в дан-

ном случае приемник сигналов с известными запаздываниями (например, приемник, реализующий соотношение (3.24) или (3.25)), в котором отсчеты выходных сигналов детекторов огибающей берутся только в моменты времени, соответствующие пикам сигналов, а вклады шума не учитываются, будет иметь более высокие характеристики, чем приемник сигналов с неизвестными запаздываниями.

3.2.1.4 Приемник ОФМ-сигналов

Как отмечалось выше, фазово-когерентные методы не рассматриваются из-за сложности когерентных приемников и из-за быстрых временных изменений фаз лучевых сигналов, принимаемых подвижной станцией. (В разделе 3.2.2 будет показано, как можно построить когерентный приемник, используя когерентный рекурсивный интегратор.) Однако разностно-когерентные методы оказываются вполне применимыми, когда на интервале между двумя последовательно переданными сигналами фазы лучей существенно не изменяются. Это — обычный для практики случай. И хотя структура оптимального разностно-когерентного приемника многолучевого сигнала нам не известна, она, по-видимому, должна быть весьма сходной со структурой приемника при однолучевом распространении сигнала (рис.3.7).

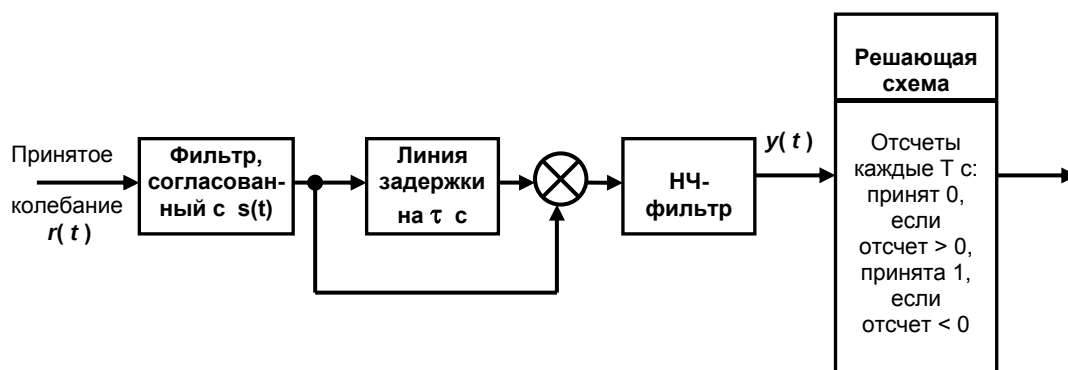


Рис. 3.7. Оптимальный приемник однолучевого ОФМ-сигнала

Предположим теперь, что в многолучевом случае, удовлетворяющем условиям упрощенной модели (распределенные по пуассоновскому закону запаздывания разрешимых лучей и независимые интенсивности), оптимальный приемник ОФМ-сигнала также связан с приемником, показанным на рис. 3.7, как и приемник, показанный на рис.3.5, с приемником, показанным на рис. 3.3. Иными словами, мы будем полагать, что этот оптимальный приемник имеет вид, показанный на рис. 3.8.

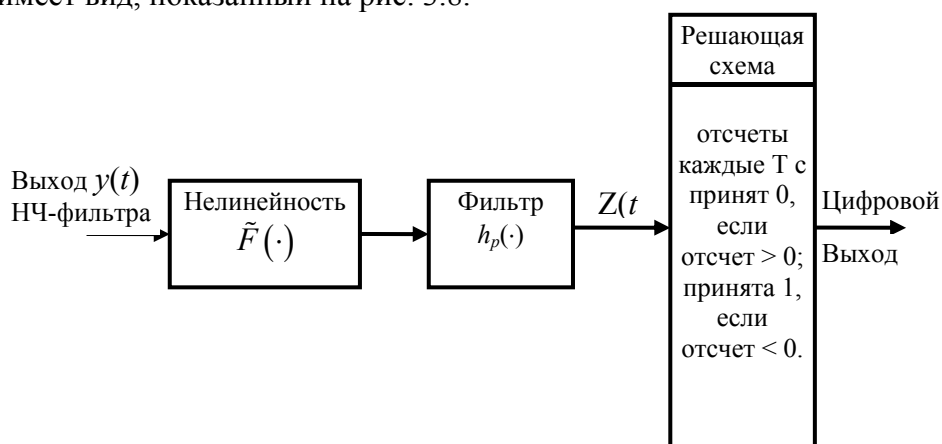


Рис. 3.8. Оптимальный приемник ОФМ-сигналов для многолучевого канала с неизвестными запаздываниями лучей.

На этом рисунке нелинейность

$$\tilde{F}(y) = F\left(\sqrt{|y|}\right) \text{sign}(y) \quad (3.28)$$

является биполярным вариантом нелинейности F , что обусловлено тем фактом, что входом этой нелинейности является квадрат огибающей выхода согласованного фильтра, а не сама огибающая, как на рис. 3.5. График функции $\tilde{F}$ для случая, когда F определяется выражением (3.28), показан на рис. 3.10 (ср. с рис. 3.6).

Решающая схема на рис. 3.8 с периодом T берет отсчеты сигнала на выходе интегрирующего фильтра $h_p(\bullet)$, с тем, чтобы определить экстремумы этого сигнала. Форма колебаний в различных точках приемника показана на рис. 3.9. В отсутствие шума выход $y(t)$ детектора произведения в схеме на рис. 3.7 и, следовательно, вход нелинейности в схеме на рис. 3.8, будут приближенно иметь вид

$$y(t) = (-1)^d \sum_{k=0}^{K-1} a_k^2 \left| \gamma(t - 2T - t_k) \right|^2, \quad 2T \leq t \leq 2T + \Delta, \quad (3.29)$$

при передаче символа $d = 0$ или 1 , принятого на интервале $[T, 2T]$.

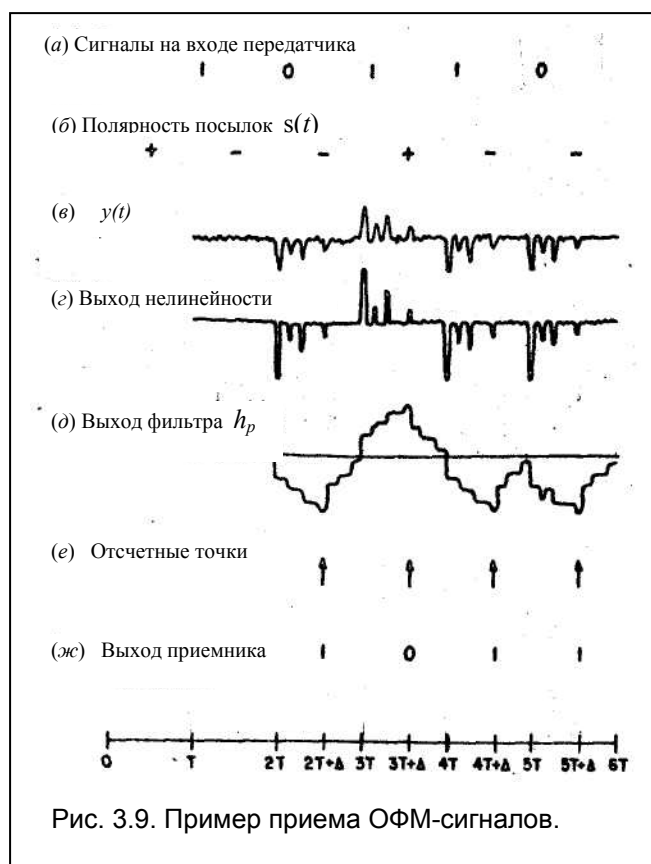


Рис. 3.9. Пример приема ОФМ-сигналов.

Пример такого «незашумленного» выхода $y(t)$ показан на рис. 3.9, в для входной двоичной последовательности 101100. Выходы нелинейности F и фильтра $h_p(\bullet)$

$$h_p(t) = \begin{cases} 1, & 0 \leq t \leq \Delta; \\ 0 & \text{иначе} \end{cases} \quad (3.29a)$$

показаны на рис. 3.9, г, д. Отсчеты выхода фильтра $h_p(\bullet)$ берутся в моменты времени $nT + \Delta$, $n=2, 3, \dots$, и используются для формирования выходных сигналов приемника, задержанных на $(T+\Delta)$ с относительно соответствующих входных символов.

3.2.1.5 Приемники, зондирующие канал

До сих пор, рассматривая структуру приемника, мы исходили из априорных сведений о статистиках канала. Однако во многих случаях характеристики канала можно измерить с помощью зондирующих сигналов, что позволяет определить апостериорные статистики. В качестве зондирующих сигналов могут использоваться либо специ-

альные сигналы, либо сами информационные сигналы; в последнем случае зондирование и передача данных осуществляются одновременно.

На первый взгляд может показаться, что для зондирования можно непосредственно использовать приемники, которые обсуждались выше при рассмотрении случая неизвестных запаздываний, а характеристики функций F и фильтра $h_p(\bullet)$ определять не по априорным, а по апостериорным статистикам. Действительно, такой подход уместен в отношении распределений интенсивностей лучей и нелинейностей F , которые они определяют.

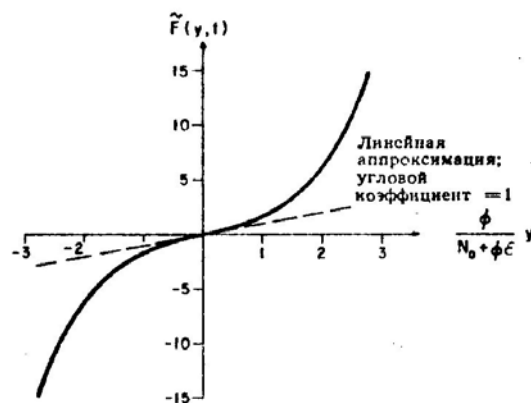


Рис. 3.10. Характеристика нелинейности

Обычно используется другой подход, в котором предполагается, что в результате зондирования получаются очень точные оценки всех трех переменных $\{\tau_k\}$, $\{a_k\}$ и $\{\theta_k\}$ или хотя бы двух первых из них. Эти оценки полагаются точными и используются в приемнике в качестве параметров.

Ниже мы ограничимся рассмотрением некогерентных приемников с зондированием канала, в которых используются только оценки интенсивностей и запаздываний лучей.

«Оптимальный» приемник M -лучевого сигнала, в котором используются оценки интенсивностей и запаздываний, можно, очевидно, реализовать на основе выражения (3.24). Такой приемник должен вычислять величины

$$w_l = \sum_{k=0}^{K-1} \ln I_0 \left(\frac{2\hat{a}_k \hat{x}_{lk}}{N_0} \right) \quad (3.30)$$

где $\hat{a}_k$ — оценка интенсивности k -го луча, $\hat{x}_{lk}$ — отсчет огибающей сигнала на выходе l -го согласованного фильтра в момент времени $t = T + t_k$, а $\hat{t}_k$ — оценка запаздывания k -го луча. Индекс $l=1, \dots, M$, для которого величина w_l максимальна, является цифровым выходом приемника. Если ОСШ для каждого луча достаточно велико, то (3.30) можно аппроксимировать выражением

$$w_l \cong \frac{2}{N_0} \sum_{k=0}^{K-1} \hat{a}_k \hat{x}_{lk}, \quad (3.31)$$

так как при больших x $\ln I_0(x) \approx x$. Выражение (3.31) описывает оптимальный линейный сумматор Бреннана [8], используемый при разнесенном приеме сигналов; мы также будем его использовать, поэтому остановимся на нем несколько подробнее.

Линейный сумматор, описываемый выражением (3.31), можно реализовать с помощью трансверсального фильтра, как показано на рис. 3.11. Этот фильтр состоит из линии задержки на время Δ , у которой сделаны отводы с интервалом, равным по крайней мере $1/W$, так что всего имеется $W\Delta$ отводов. На вход трансверсального фильтра подается напряжение с выхода детектора огибающей, например $X_{ji}(t)$ (см. рис. 3.5). Выход трансверсального фильтра представляет собой взвешенную сумму сигналов с определенных отводов, номера которых зависят от оценок запаздываний лучей.

Оценки t_k ($k=0, \dots, K-1$) используются для включения усилителей, подсоединенных к отводам, задержки которых (измеряемые с правого конца линии) наиболее точно аппроксимируют оценки $\hat{t}_k$. Иными словами, активируется K усилителей, коэффициенты усиления которых устанавливаются пропорционально величинам соответствующих оценок интенсивностей $\hat{a}_k$. На рис. 3.11 указаны коэффициенты усиления усилителей при приеме четырех лучевого сигнала.

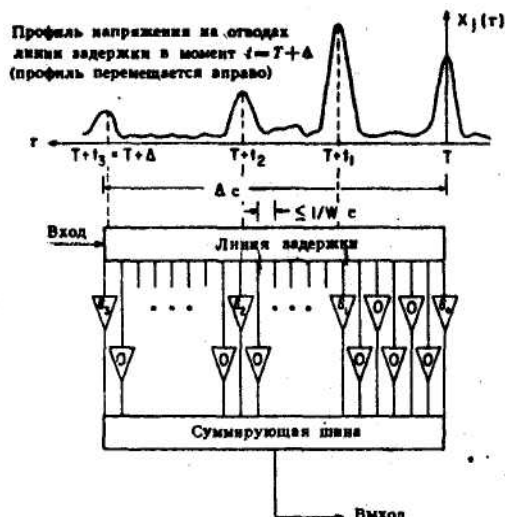


Рис. 3.11. Трансверсальный фильтр.

Для того чтобы лучше понять, как работает трансверсальный фильтр, на рис. 3.11 показан профиль изменения напряжения вдоль линии задержки для случая, когда на ее вход поступает сигнал $x_j(t)$ с выхода детектора огибающей, включенного после j -го согласованного фильтра (см. рис. 3.4). Этот профиль постепенно сдвигается вправо, и на рис. 3.11 его положение соответствует моменту времени $t = T + \Delta$. В этот момент пики сигнала соответствуют задержкам $t_0, t_1, \dots, t_{K-1}$, измеряемым с правого конца линии. На рис. 3.11 изображен случай, когда $t_0 = 0$ и $K=4$, причем предполагается, что $t_s = \Delta$, где Δ — мак-

симальное относительное запаздывание (т. е. превышение над t_0 , ожидаемое в данном канале). Если $\hat{t}_k$ — точные оценки времен запаздывания, то в момент времени $t=T+\Delta$, напряжение на выходе трансверсального фильтра будет приближенно равно величине:

$$\sum_{k=0}^{K-1} \hat{a}_k \hat{x}_j(T + \hat{t}_k) \cong \sum_{k=0}^{K-1} \hat{a}_k \hat{x}_{jk}, \quad (3.22)$$

т. е. оказывается пропорциональной величине w_l , определяемой (3.21) при $l = j$.

Таким образом, трансверсальный фильтр является фильтром, согласованным с оценкой значения низкочастотного эквивалента импульсной характеристики канала $h_c(t)$.

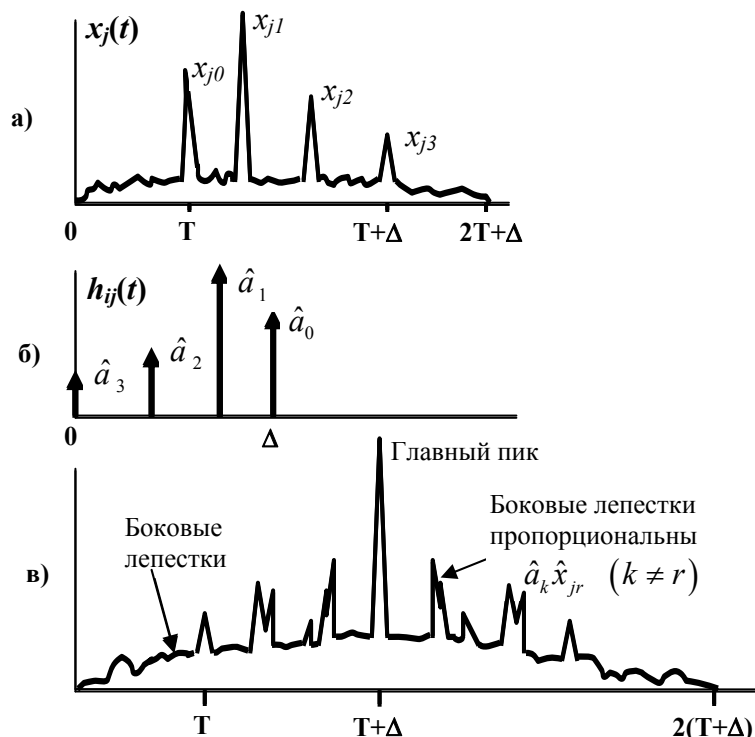


Рис. 3.12. (а) – вход фильтра; (б) – импульсная характеристика фильтра; (в) – Выход фильтра (свертка входа и импульсной характеристики фильтра)

В процессе перемещения $x_j(t)$ на рис. 3.11 вправо трансверсальный фильтр осуществляет свертку $x_j(t)$ с $h_{ij}(t)$, показанных соответственно на рис. 3.12, а и 3.12, б, в результате чего формируется выходной сигнал, показанный на рис. 3.12, в. Главный пик этого сигнала соответствует моменту времени $t=T+\Delta$, т. е. тому моменту, когда входной сигнал $x_j(t)$ занял такое положение, при котором его пики расположены точно над активированными отводами линии задержки. Высота главного пика пропорциональна величине (3.22). Выходной сигнал имеет также небольшие побочные пики, пропорциональные величинам $\hat{a}_k \hat{x}_{jr}$ ($k=0, 1, 2, 3; r \neq k$), и боковые лепестки, накопленные в результате свертки $h_{ij}(t)$ с боковыми лепестками сигнала $x_j(t)$. Ясно, что для получения (3.22) необходимо оставить только один главный пик.

Таким образом, полный приемник M -лучевого сигнала, в котором используются оценки характеристик канала, будет иметь вид, показанный на рис. 3.13. Выход $x_l(t)$, $l=1, \dots, M$ каждого детектора огибающей подается на вход трансверсального фильтра типа показанного на рис. 3.11, параметры которого изменяются в соответствии с оценками $\{\hat{t}_k\}$ и $\{\hat{a}_k\}$, полученными от зондирующего приемника. В момент времени $t= T+ \Delta$ выходы трансверсальных фильтров считываются и сравниваются, в результате чего определяется индекс наибольшего из них. Такой приемник получил название «Рейк» из-за внешнего сходства профиля напряжения на отводах трансверсального фильтра с зубьями гребня или грабель [гаке (рейк) — грабли].

Зондирующий приемник, показанный на рис. 3.13, также может иметь структуру, образованную последовательным соединением согласованного фильтра, детектора огибающей и линии задержки с отводами, причем согласованный фильтр согласуется с известным колебанием, которое может иметь большую энергию и длительность, чем информационные сигналы s_j . Перед передачей данных (или во время передач, если характеристики канала изменяются достаточно быстро) зондирующий приемник «слушает» свой сигнал. Когда сигнал принят, огибающая напряжения на выходе согласованного фильтра будет иметь множество пиков, т. е. будет иметь вид, аналогичный показанному на рис. 3.12, а. Затем это колебание будет перемещаться по линии задержки так же, как показано на рис. 3.11. Когда первый пик сигнала

достигает правого конца линии, срабатывает пороговый элемент и берутся отсчеты напряжений на всех отводах линии задержки, и их значения, которые пропорциональны интенсивностям лучей, запоминаются, если они превышают некоторую пороговую величину, в противном случае они полагаются равными нулю. Далее эти отсчеты используются для установки коэффициентов усиления усилителей, подключенных к отводам трансверсальных фильтров приемника информационных сигналов.

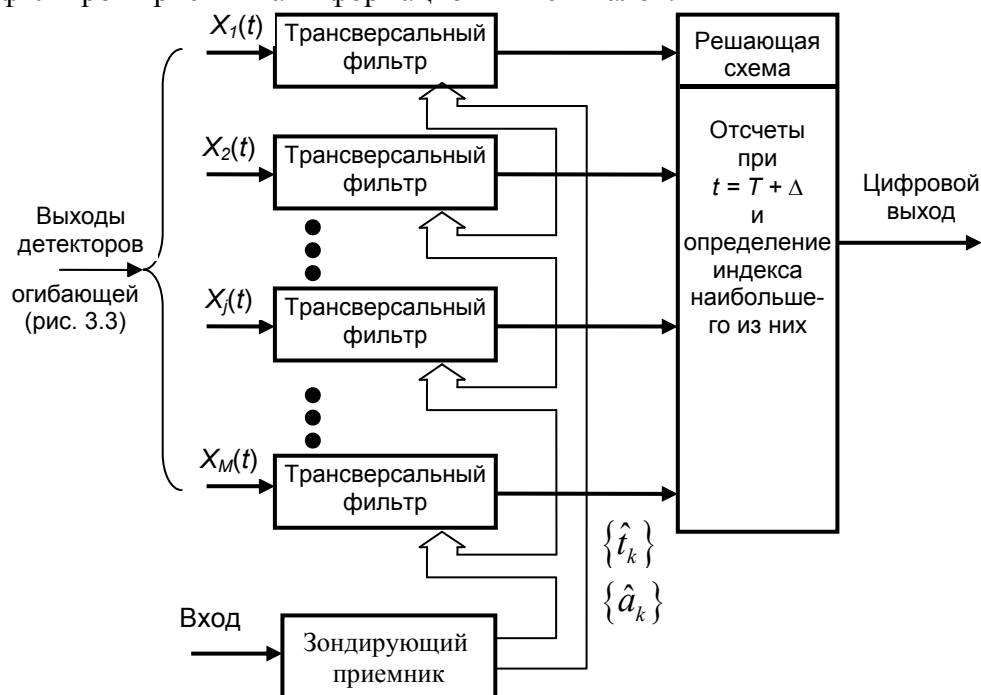


Рис. 3.13. Приемник многолучевого сигнала с зондированием канала (приемник системы «Рейк»).

В самонастраивающемся варианте этого приемника для зондирования канала используются сами информационные сигналы. В момент времени, когда принято решение о том, что принят, скажем, j -й сигнал, напряжения на отводах j -го трансверсального фильтра фиксируются точно так же, как это было описано выше для зондирующего приемника. Затем значения этих напряжений используются для установки коэффициентов усиления соответствующих усилителей во всех трансверсальных фильтрах.

Следует отметить, что механизм зондирования выполняет также и функцию синхронизации приемника, которую мы пока не рассматривали. И в приемнике с отдельным зондированием, и в самонастраивающемся приемнике при срабатывании порогового устройства, подключенного к крайнему правому отводу линии задержки, запускаются часы (генератор синхронизирующих импульсов), с помощью которых обеспечивается соответствующая синхронизация отсчетных импульсов, поступающих на входы решающей схемы.

Наконец, отметим, что приемник ОФМ-сигнала можно построить, используя те же принципы, которые были заложены в приемнике системы «Рейк».

Комбинирование сигналов на основе RAKE-приемника реализовано в стандарта CDMA-2000. Приемные устройства мобильной станции (МС) и базовой станции (БС) включают несколько (3 на МС и 4 на БС) параллельно работающих корреляторов, которые выделяют наиболее сильные сигналы. Выходы корреляторов сводятся к одному и тому же моменту времени и суммируются. Тем самым эффект многолучевого распространения используется для повышения качества связи.

3.2.1.6 Межсимвольная интерференция

До сих пор мы пренебрегали эффектами межсимвольной интерференции, поскольку основное внимание уделялось приемнику отдельных изолированных посылок (см., например, рис. 3.4). Даже в случае ОФМ-сигнала, когда информация передается посредством изменения или сохранения полярностей двух последовательных посылок, эти посылки полагались достаточно длинными или достаточно разнесенными по времени, благодаря чему временная дисперсия, обусловленная многолучевостью канала, приводила лишь к незначительному ухудшению рабочих характеристик или же вообще на них не влияла (см. рис. 3.12).

Предположим теперь, что каждая посылка имеет длительность T , а последовательная передача любых посылок осуществляется с интервалом T_s . В соответствии с (3.23) отклики согласованных фильтров в рассмотренных выше приемниках на каждую посылку длиной T будут длиться $2T+\Delta$, что может привести к межсимвольной интерференции.

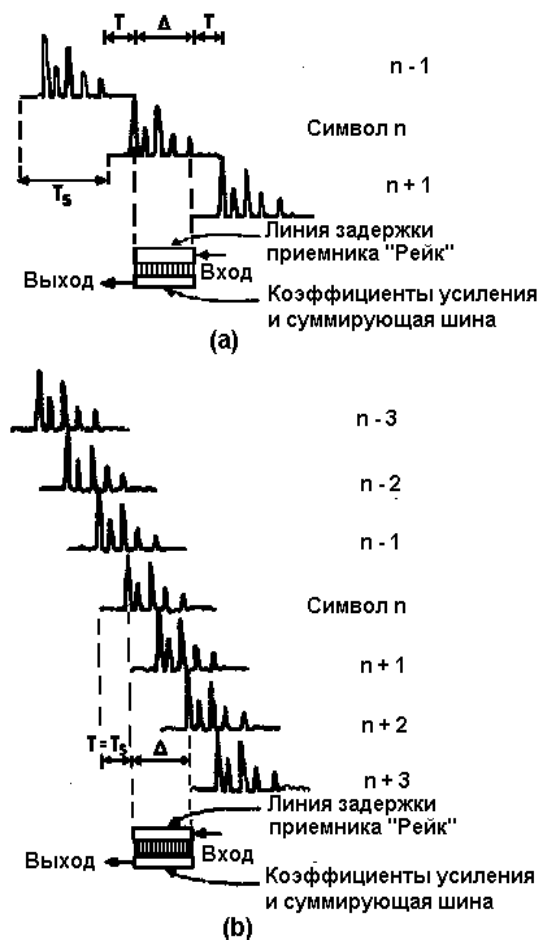


Рис.3.14. а) Межсимвольная интерференция отсутствует; б) $T_s=T$; $\Delta=2T$ — умеренная межсимвольная интерференция

точно соответствуют отводам линии задержки. Однако в данном случае два предшествующих ($(n-2)$ -й и $(n-1)$ -й) сигнала еще частично не вышли из линии задержки, а два последующих ($(n+1)$ -й и $(n+2)$ -й) уже частично вошли в нее, и это обстоятельство приводит к возникновению межсимвольной интерференции. В общем случае с каждым символом будет интерферировать $2[\Delta/T]$ предшествующих и последующих символов, где $[x]$ — наименьшее целое число, большее или равное x .

Для того чтобы полностью устранить межсимвольную интерференцию, необходимо выполнить условие $T_s \geq T + \Delta$. Поскольку величина Δ для данного канала фиксирована, то остается лишь несколько уменьшить величину T (однако выбор T зависит от требования $TW \gg 1$ и ограничений, налагаемых на ширину полосы W). Следовательно, в бинарных системах мы не можем полностью устранить межсимвольную интерференцию, если скорость

Для того чтобы наглядно проиллюстрировать возникновение межсимвольной интерференции в широкополосном случае ($TW \gg 1$), на рис. 3.14 показана последовательность огибающих откликов отдельного согласованного фильтра на периодически повторяемый входной сигнал при двух значениях T_s .

На рис. 3.14 а показаны огибающие откликов согласованного фильтра на три последовательные посылки для случая, когда $T_s \geq T + \Delta$. Для наглядности показаны также линия задержки приемника «Рейк» длиной Δ с и сумматор (см. рис. 3.11). Межсимвольная интерференция в данном случае полностью отсутствует. Указанная последовательность огибающих выходных напряжений фильтра «заморожена» в тот момент времени, когда пики центрального (n -го) многолучевого входного сигнала точно соответствуют отводам линии задержки и «готовы» для взятия отсчетов. Заметим, что в данный момент $(n-1)$ -й и $(n+1)$ -й отклики «находятся» вне линии задержки, и, следовательно, не могут интерферировать с n -м откликом.

На рис. 3.14 б показан случай умеренной межсимвольной интерференции. Здесь $T_s = T$, а $\Delta = 2T$, так что $T_s = (T + \Delta)/3$. Огибающие откликов показаны для семи последовательных посылок, которые зафиксированы в тот момент времени, когда пики n -го многолучевого входного сигнала

передачи данных превышает величину, равную примерно $1/\Delta$ бит/с. Так, например, при работе такой системы в городских условиях (эффективное значение $\Delta=5$ мкс) со скоростью, большей примерно 200 кбит/с, межсимвольная интерференция будет возникать даже в том случае, когда используются широкополосные сигналы.

Используемый в данном случае подход основан на том факте, что, хотя в момент принятия решения о принятом символе, многолучевые сигналы, соответствующие предыдущему и последующему символам, частично находятся в линии задержки (см. рис. 3.14 б), вероятность того, что их пики будут точно соответствовать активированным отводам линии, весьма мала. Напомним, что отводы у линии задержки сделаны через каждые $1/W$ с. Так как $TW \gg 1$ и так как мы теперь полагаем $\Delta/T > 1$, то $W\Delta \gg 1$ и это означает, что линия задержки должна иметь большое число отводов. Например, при $TW=100$ и $\Delta/T=5$ линия задержки будет иметь $W\Delta=500$ отводов. Однако при приеме сигнала активируются только те отводы, на которых ожидаются пики этого сигнала, поэтому в рассмотренном выше примере даже при наличии, скажем, 20 лучей будет активироваться всего $20/500 = 2,5\%$ от общего числа отводов. Эти отводы

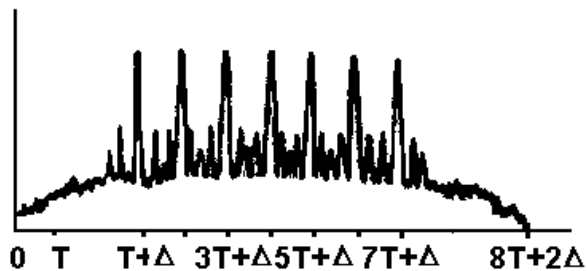


Рис.3.15. Реакция ТФ приемника «Рейк» на 7 сигналов рис. 3.14.

будут соответствовать пикам центрального n -го отклика на рис. 3.14 б; при этом весьма маловероятно, что пики смежных с ним откликов также попадут на эти активированные отводы.

Последовательность импульсов после семи таких сверток и соответствующего их объединения будет иметь примерно такой вид, как показано на рис. 3.15. Каждый из главных пиков соответствует здесь точному согласованию одной последовательности импульсов с активированными

отводами линии задержки, и именно по ним берутся далее отсчеты. Пьедестал, на котором расположены эти пики, состоит из таких же небольших побочных пиков, как и на рис. 3.12, и смеси боковых лепестков. Этот пьедестал мы будем называть «помехой, обусловленной многолучевостью». Способность приемника системы «Рейк» выделять именно ту часть отклика согласованного фильтра, которая соответствует принимаемому в данный момент символу, обусловлена структурой применяемой в нем линии задержки с отводами.

Суммируя изложенное выше, нетрудно видеть, что приемники системы «Рейк» могут работать со значительно более высокими, чем $1/\Delta$ бит/с, скоростями передачи данных без заметного ухудшения характеристик.

Ниже будут рассмотрены другие приемники многолучевых ОФМ-сигналов, структура которых проще приемника «Рейк».

3.2.1.7 Приемник с последетекторным интегрированием сигналов

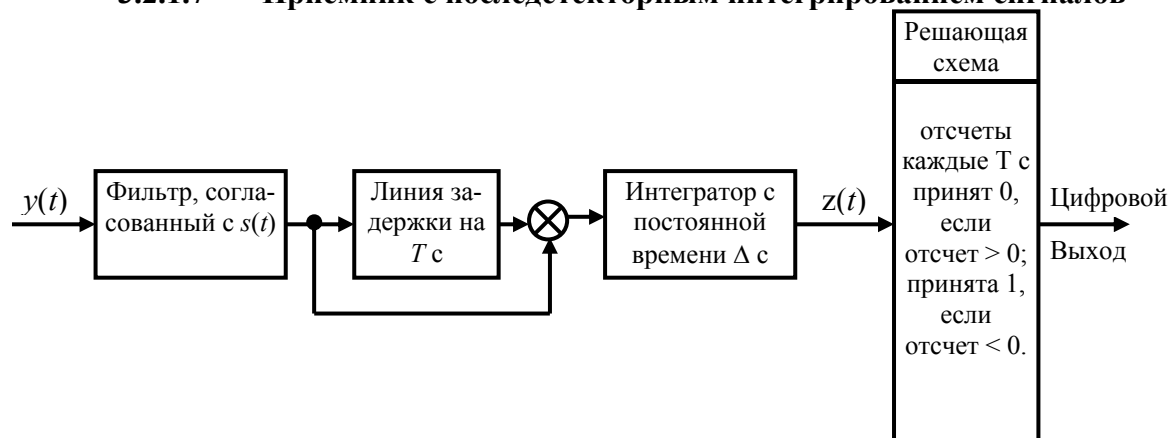


Рис. 3.16. ПДИ – квазиоптимальный приемник ОФМ-сигнала для неизвестных времен запаздывания лучей.

Роль фильтра, согласованного с каналом, в приемнике с последетекторным интегрированием (ПДИ-приемник, или просто ПДИ), представленным рис. 3.16, играет интегратор с постоянной времени интегрирования Δ , импульсная характеристика которого имеет вид (3.29a). Эта характеристика предполагает, что запаздывания лучей равномерно распределены на интервале $[0, \Delta]$ и $p(t_k) = 1/\Delta$, $0 \leq t_k \leq \Delta$. В действительности запаздывания лучей распределены неравномерно, поэтому характеристика $h_p(t)$, определяемая выражением (3.29a), будет субоптимальной. Так как НЧ-фильтр на рис. 3.7 необходим лишь для устранения составляющих удвоенной частоты на выходе умножителя, то на рис. 3.16 его функция возложена на интегрирующий фильтр, который также является фильтром низких частот.

Вход решающей схемы на рис. 3.16 будет все еще иметь вид, сходный с показанным на рис. 3.9 д, но теперь он будет получаться в результате интегрирования напряжения $y(t)$ (рис. 3.9 в), а не напряжения, показанного на рис. 3.9 г. Отсчеты выхода интегрирующего фильтра будут, как и прежде, браться в моменты времени, указанные на рис. 3.9 е.

Расчет характеристик ПДИ проводился в предположении, что:

- передаваемые сигналы ортогональны и имеют нулевые боковые лепестки;
- нелинейность на рис. 3.8 имеет квадратичный характер;
- все лучи независимы и их интенсивности имеют одинаковые релейевские распределения;
- все лучи разрешимы;
- $\Delta \ll T$, МСИ отсутствует;
- в канале присутствует аддитивный белый шум с односторонней плотностью мощности N_0 .

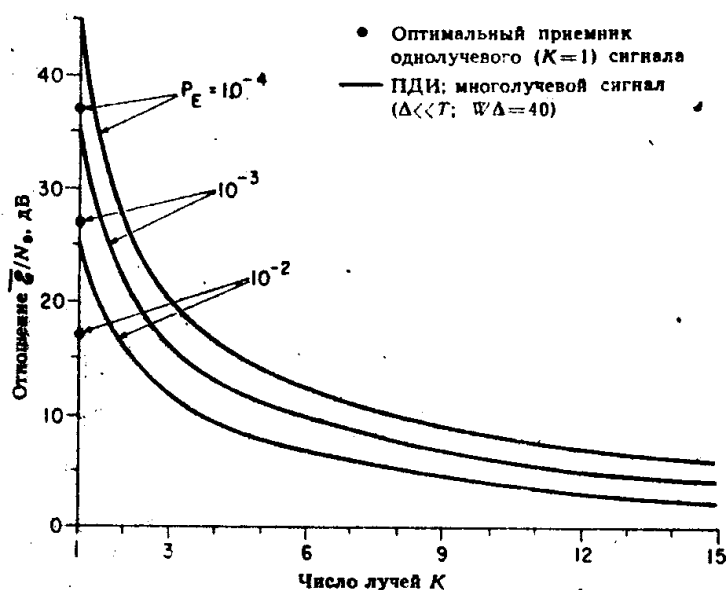


Рис. 3.17 Зависимость среднего на 1 луч отношения $\bar{S}/N_0$ для ОФМ-приемника с последетекторным интегрированием сигнала, требуемого для обеспечения заданной вероятности ошибки на 1 бит, от числа лучей K (независимые лучи с релейевскими замираниями).

Результаты расчета для $W\Delta = 40$ представлены на рис. 3.17.

Поскольку ранее принималось, что для ПДИ $\Delta \ll T$, то это, в свою очередь означает, что коэффициент расширения спектра сигнала $TW \gg 40$.

Для сравнения на рис. 3.17 указаны значения вероятности ошибки для оптимального однолучевого ОФМ-приемника. Нетрудно видеть, что при наличии ряда релейевских лучей с близкими среднеквадратичными интенсивностями, ПДИ значительно (на 5 – 20 дБ) превосходит приемник, который принимает сигнал по одному лучу (например, по ЛВЗ) и пренебрегает другими лучами. Это сравнение весьма убедительно иллюстрирует

улучшение характеристик при разнесенном приеме, которое может обеспечить многолучевой канал.

3.2.1.8 Цифровой вариант приемника «Рейк»

В приемниках с зондированием канала (рис. 3.13 и 17) установка коэффициентов усиления усилителей, подключенных к отводам трансверсального фильтра (или фильтров) осуществляется в соответствии с оценками интенсивностей лучей. Однако для реализации значительно проще приемник (правда, он будет иметь несколько худшие характеристики), в котором отводы

трансверсального фильтра просто подключаются к суммирующей шине в соответствии с текущими оценками времен запаздывания лучей. В этом случае необходимость получения оценок интенсивностей лучей отпадает, и зондирующий приемник должен лишь определять оценки времен запаздывания лучей $\{\hat{t}_k\}$. Такой упрощенный вариант приемника мы будем называть «цифровым Рейк». Структура этого приемника при использовании ОФМ-сигналов показана на рис. 3.18, где в целях дальнейшего упрощения исключен также и вычислитель квадратного корня.

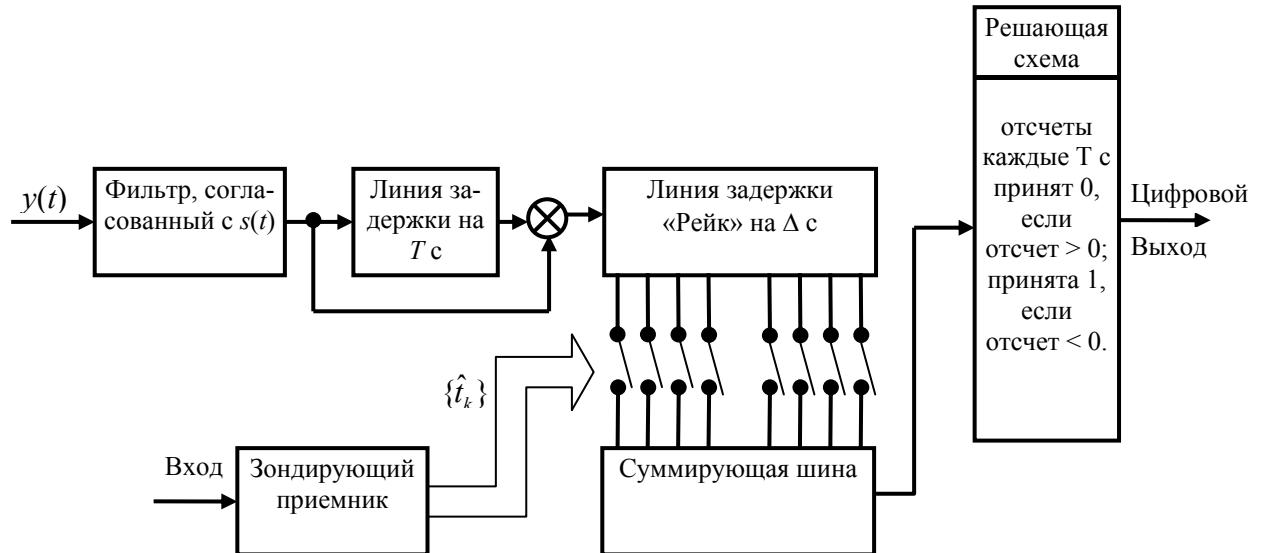


Рис. 3.18. Цифровой вариант приемника «Рейк» – квазиоптимальный приемник ОФМ-сигнала, в котором используются оценки времен запаздывания лучей.

Оценки запаздываний лучей, поступающие от зондирующего приемника, используются здесь для замыкания соответствующих ключей в цепях отводов трансверсального фильтра, в результате чего эти отводы подключаются к суммирующей шине. Заметим, что при замыкании всех ключей приемник, показанный из рис. 3.18, оказывается примерно аналогичным ФМ-приемнику, показанному на рис. 3.16, у которого операция интегрирования на интервале Δc

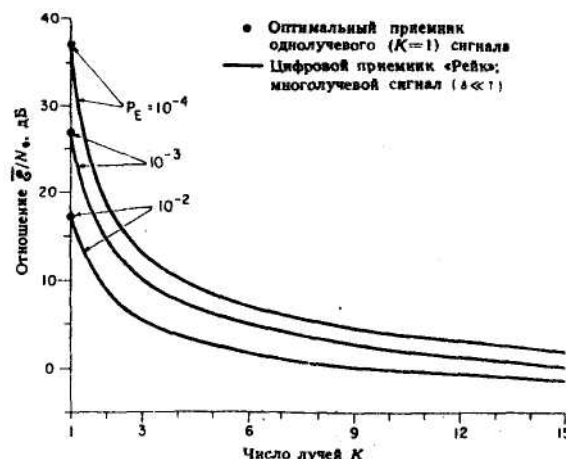


Рис. 3.19. Зависимость среднего на 1 луч ОСШ для цифрового Рейк от числа лучей, требуемого для обеспечения заданной вероятности ошибки на 1 бит.

заменена операцией суммирования в дискретных точках этого интервала. Характеристики этого приемника, аналогичные кривым рис. 3.17, представлены на рис. 3.19 для случая релейских замираний без учета межлучевой и межсимвольной интерференции, наличие которой можно учесть введением дополнительного эквивалентного шума.

3.2.1.9 Сравнение различных приемников ОФМ-сигналов

Сигнал, использованный в приведенных ниже экспериментах, имеет следующие параметры: ширина полосы $W = 1/100$ нс ≈ 10 МГц; коэффициент расширения спектра $TW = 127$, скорость передачи $1/T = 78,7$ кбит/с. Кроме того, поскольку на каждом профиле многолучевости брался 71 отсчет с интервалом 100 нс, то $W\Delta = 70$. Таким образом, $\Delta/T = 0,55$, и, следовательно, межсимвольной интерференцией можно пренебречь.

При обработке принятого ОФМ-сигнала использовалось семь различных процедур принятия решения, соответствующих семи типам приемника.

- 1) *Первый луч (ПЛ)*. Здесь использовался классический механизм, при котором фиксируется первое пересечение порога напряжением $|y(t)|$, и решение о принятом символе (0 или 1) принимается в соответствии со знаком $y(t)$.
- 2) *Луч наибольшей интенсивности (ЛНИ)*. Здесь на интервале передачи каждого бита определялся наибольший пик сигнала $1 \# (/) 1$, и решение принималось в зависимости от его полярности. И снова для некоторого упрощения было принято, что наибольший пик обусловлен наличием сигнала. Момент появления наибольшего пика фиксировался как момент прихода луча наибольшей интенсивности a_k .
- 3) *Последетекторный интегратор (ПДИ)*. Для этого случая механизм принятия решения указан на рис. 3.16. Однако интегрирование осуществлялось не в полном интервале запаздываний лучей, а в интервале Δ , который только на 4 мкс превышает величину ЛВЗ.
- 4) *Адаптивный последетекторный интегратор (АПДИ)*. Здесь ПДИ был сделан адаптивным благодаря тому, что интегрирование выполнялось только на интервале между первым и последним пересечениями порога сигналом $|y(t)|$. Пороговый уровень и метод определения моментов первого и последнего пересечений идентичны использованному при определении первого луча (ПЛ).
- 5) *Взвешенный последетекторный интегратор (ВПДИ)*. Установив тот факт, что последний луч в профиле многолучевости является, как правило, более слабым и менее частым, чем предыдущие лучи, прямоугольную импульсную характеристику $h_p(t)$, определяемую выражением (3.29а), заменили экспоненциальной весовой функцией $h_p(t) = \exp[-(\Delta-t)/\tau]$, $0 \leq t \leq T$, где постоянная времени τ была принята равной 1,5 мкс.
- 6) *Цифровой приемник «Рейк» (Ц-Рейк)*. Структура этого приемника показана на рис. 3.18. Положения ключей в этой схеме определялись в соответствии с предположением, что в качестве лучей зондирующий приемник будет идентифицировать лишь те из них, интенсивность которых достаточна для превышения сигналом $|y(t)|$ порога, установленного в информационном приемнике. Пороговый уровень и метод определения лучей, превышающих этот уровень, идентичны использованным при определении первого луча (ПЛ).
- 7) Приемник «Рейк» - описан выше, рис. 3.13.

Расчетные кривые вероятности ошибки для перечисленных типов приемников представлены

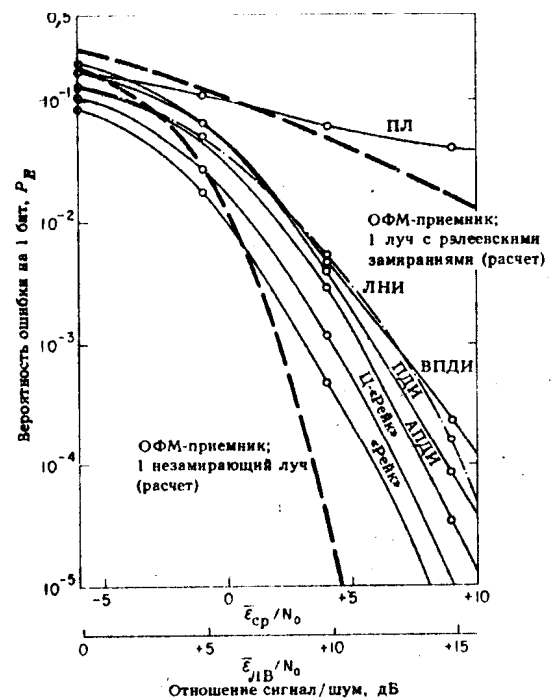


Рис. 3.20. Зависимости вероятности ошибки на 1 бит от величины ОСШ для различных приемников в условиях многолучевого распространения радиоволн в районе с плотной многоэтажной застройкой [xxx]/

на рис. 3.20.

Сравнение кривых для ВПДИ и ПДИ показывает, что 4-мкс интегратор с прямоугольной весовой функцией обеспечивает лучшие результаты, чем 1,5-мкс интегратор с экспоненциальной весовой функцией по крайней мере при многолучевом распространении сигнала в районе с плотной высокоэтажной застройкой. Однако нет причин полагать, что этот вывод будет справедливым для других городских районов и других постоянных времени, результат для которых может оказаться обратным.

«ЛНИ-приемник», т. е. приемник, который «захватывает» наибольший пик в $y(t)$, работает значительно лучше «ПЛ-приемника», который «захватывает» первый из имеющихся пиков. (Это известный для разнесенного приема эффект выбора «наибольшего» сигнала.)

Выигрыш в отношении сигнал/шум при последовательном переходе от ПДИ к более сложным приемникам (АПДИ, цифровой и основной вариант приемника «Рейк») каждый раз составляет примерно 1 дБ при условии, что во всех этих приемниках определяются точные оценки параметров лучей. Столь умеренное улучшение характеристик у каждой последующей системы подтверждает наш вывод об их субоптимальности.

Переход от ПДИ (случай неизвестных запаздываний) к цифровому приемнику «Рейк» (случай известных запаздываний) дает выигрыш, равный 2 дБ, что, по-видимому, говорит о том, что в случае неизвестных запаздываний ПДИ весьма близок к оптимальному приемнику.

Переход от цифрового варианта приемника «Рейк» (случай неизвестных интенсивностей и известных запаздываний) к основной схеме приемника «Рейк» (случай известных интенсивностей и запаздываний) дает выигрыш всего лишь в 1 дБ, откуда можно заключить, что цифровой приемник «Рейк» близок к оптимальному в случае, когда зондирующий приемник дает лишь оценки запаздываний лучей, а приемник «Рейк» — в случае, когда зондирующий приемник дает оценки интенсивностей и запаздываний лучей.

Вследствие того что в ВПДИ, ПДИ, АПДИ и в цифровом и основном вариантах приемника «Рейк» используется энергия большого числа почти независимых лучей, эти приемники обладают значительным преимуществом перед приемником, в котором используется энергия только одного подверженного замираниям луча. При малых отношениях сигнал/шум их характеристики даже несколько превосходят характеристики приемника, использующего энергию одного незамирающего луча. При больших отношениях сигнал/шум их преимущество теряется, однако и в этом случае их характеристики всего лишь на несколько децибел уступают характеристикам однолучевого приемника при отсутствии замираний.

Если помимо характеристик необходимо также учитывать сложность аппаратуры, то при невысоких скоростях передачи данных наиболее приемлемым будет, по-видимому, ПДИ-приемник, значительно более простая реализация которого по сравнению с цифровым и основным вариантами приемника «Рейк» «покупается» за счет проигрыша в отношении сигнал/шум, равного соответственно 2 и 3 дБ. АПДИ, реализация которого также относительно проста, превосходит ПДИ менее чем на 1 дБ, а это, безусловно, недостаточная плата за введение дополнительного порога, АРУ и запоминающих элементов.

3.2.2 Установка пакетной радиосвязи (УПР)

Пакетные радиосети (ПРС), начавшие развиваться в 1960-е годы, получили широкое распространение в коммерческой, военной и научной областях. В 1972 году Управление перспективных научно-исследовательских работ Министерства обороны США (DARPA) – то самое агентство, которое стояло у истоков Internet – приступило к проведению исследований с целью создания собственной пакетной радиосети. В их задачу входил поиск эффективных методов коллективного использования выделенного радиоканала в условиях неполной или непостоянной связности, характерной для связи с подвижными объектами. В ходе разработки потребовалось решить целый ряд технических проблем, начиная с объединения высокоскоростных модемов и высокопроизводительных микропроцессоров в малогабаритном недорогом блоке с целью реализации протоколов коллективного использования канала и кончая созданием алгоритмов маршрутизации пакетов в сетях с быстроменяющимися характеристиками.

Ниже будет рассмотрен приемопередатчик УПР [9], созданной в рамках этого проекта, а точнее – обработку в нем принимаемых широкополосных многолучевых сигналов. Во второй главе было показано, что когерентный прием сигналов с абсолютной фазовой манипуляцией наталкивается на большие, практически непреодолимые, трудности, связанные с формированием опорного колебания, фаза которого однозначно связана с фазой принимаемого сигнала. Поэтому в контексте настоящего пособия весьма уместно познакомиться с решением этой проблемы, найденной в [9].

Не касаясь других, подчас довольно интересных, технических решений и характеристик указанной УПР, детально описанных в [9], рассмотрим демодулятор этой установки (рис. 3.21).

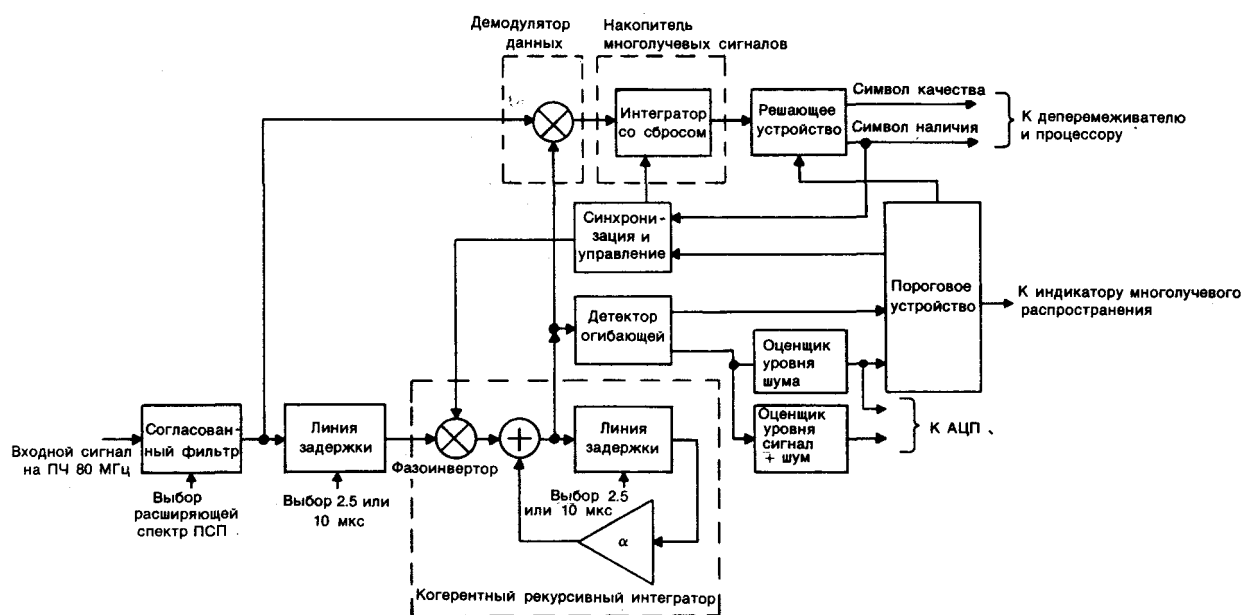


Рис. 3.21. Функциональная схема демодулятора УПР.

Приведем некоторые характеристики УПР, необходимые для понимания работы демодулятора. УПР использует классические сигналы БФМ, спектр которых расширен кодовой последовательностью (ПСП) с тактовой частотой 12,8 МГц. Кодовая последовательность может меняться от символа к символу; информация о текущем коде содержится в преамбуле пакета. Передача осуществляется на двух скоростях: 400 Кбит/с (высокая, длительность символа 2,5 мкс) и 100 Кбит/с (низкая, длительность символа 10 мкс). Приемник с двойным преобразованием частоты имеет вторую ПЧ, равную 80 МГц.

В демодуляторе с помощью программируемого согласованного фильтра (СФ) снимают модуляцию элементарными символами ПСП и получают сжатый импульс с фазовой манипуляцией информационными символами пакета (преамбула, заголовок, данные и т. д.).

1) *Согласованный фильтр НУПР*. Структурная схема реализованного программируемого согласованного фильтра НУПР приведена на рис. 3.22. Входной фильтр для приема многоэлементной ПСП, согласованный с длительностью элементарного символа, соединен с многоотводной линией задержки, на соседних выходах которой сигналы появляются с задержкой, равной длительности элемента ПСП (78.125 нс). Для формирования выходного сигнала СФ сигналы, поступающие с отводов линии задержки,

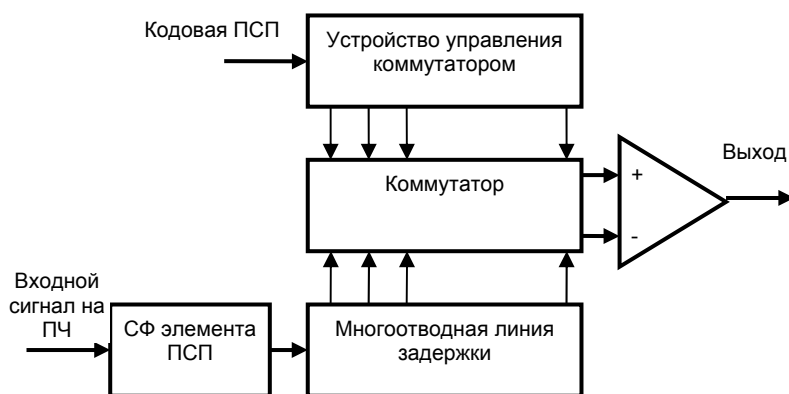


Рис. 3.22. Программируемый согласованный фильтр

суммируют в шинах с весами +1 или —1 в соответствии со структурой ожидаемой кодовой ПСП, а выходы двух шин вычитаются. Многоотводная линия задержки согласованного фильтра реализована с использованием техники ПАВ на четырех интегральных схемах, каждая из которых имеет 32 отвода. При скорости передачи данных 100 тыс. символов в секунду задействованы

все 128 отводов линии задержки и выходы четырех интегральных схем объединяются вместе. При скорости передачи данных 400 тыс. символов в секунду обрабатываются сигналы с 32 отводов линии задержки и используется выходной сигнал лишь первой интегральной схемы. Выходной сигнал СФ поступает на демодулятор данных и — после задержки на длительность одного символа — на когерентный рекурсивный интегратор (КРИ) с решающей обратной связью.

2) *Когерентный рекурсивный интегратор УПР*. КРИ обеспечивает формирование опорного колебания несущей частоты и сигналы профиля многолучевости (зависящие от структуры принимаемого многолучевого сигнала). Опорное колебание необходимо для когерентной демодуляции принимаемого радиосигнала в демодуляторе данных и после детектирования огибающей для определения факта захвата сигнала. Требуемое качество выполнения этих функций обеспечивается за счет увеличения ОСШ на выходе КРИ по сравнению с ОСШ на его входе при наличии широкополосных помех. Увеличение ОСШ зависит от коэффициента усиления α контура КРИ (рис. 3) и равно $(1+\alpha)/(1-\alpha)$. При $\alpha=0.85$ ОСШ повышается на 10.9 дБ.

3) *Накопление многолучевых сигналов и демодуляция данных*. Процессор накопления многолучевых сигналов содержит КРИ, демодулятор данных, интегратор со сбросом и следующее за ним устройство принятия решения об информационном символе. Линия задержки перед фазоинвертором КРИ необходима в связи с тем, что сигнал решения формируется в конце принятого символа. Выходной сигнал КРИ соответствует профилю многолучевости канала на радиочастоте (с точностью до фазы) и используется как когерентное опорное колебание для демодуляции данных. В демодуляторе данных происходит умножение этого опорного колебания на выходной сигнал СФ. Применение сигнала с расширенным спектром в условиях многолучевого распространения радиоволн позволяет выделить с помощью СФ сигналы отдельных лучей, взаимная задержка прихода которых приблизительно равна длительности одного элемента ПСП или около 78 нс. Взвешенные сигналы отдельных лучей суммируются в интеграторе со сбросом. Выполнение операций перемножения и интегрирования позволяет реализовать почти оптимальный коррелятор для приема входного сигнала, искаженного многолучевостью. Такая обработка сигнала обеспечивает повышение помехоустойчивости в условиях замираний при многолучевом распро-

странении радиоволн и взаимной задержки прихода сигналов отдельных лучей до 6 мкс при скорости передачи данных 100 тыс. символов в секунду и до 1.6 мкс при скорости передачи данных 400 тыс. символов в секунду.

В каждый момент принятия решения с выхода решающего устройства на вход микропроцессора УПР поступают два двоичных символа: символ наличия (т. е. решение о переданном сигнале) и символ качества. Символ наличия также поступает по цепи обратной связи на вход фазового инвертора (перемножителя) КРИ, в результате на выходе инвертора получают немодулированный импульсный сигнал, который подается на вход интегратора. Символ качества несет информацию об амплитуде принятого информационного символа и используется при работе в режиме последовательного декодирования с мягким (многоальтернативным) решением. Формируемый устройством оценки шума показатель уровня шума служит для выбора порога при формировании символа качества, значение которого используется сетью.

Таким образом, КРИ формирует опорное напряжение для фазового детектора демодулятора данных в виде текущего радиочастотного профиля многолучевости. Темп обновления опорного сигнала (профиля многолучевости) зависит от коэффициента передачи цепи обратной связи КРИ α . Так, при $\alpha = 0,85$ уровень предыдущего профиля многолучевости спадает до $\sim 0,2$ за 10 принятых символов. Демодулятор УПР обеспечивает почти оптимальный прием многолучевого сигнала БФМ; потери возникают из-за остаточной зашумленности опорного сигнала, но они невелики при выигрыше 10,9 дБ в схеме КРИ.

Ошибка в приеме символа не влечет за собой фатальных последствий, а лишь уменьшает уровень накопленного профиля многолучевости. Критическим является пакет ошибок кратностью 10 и более, но благодаря перемежению символов в передатчике и деперемежению в приемнике этой ситуации удастся избежать.

4 Системы мобильной связи

Стратегической целью развития современных общедоступных телекоммуникаций является предоставление высококачественной связи любому потребителю в любое время в любой точке земного шара. В соответствии с этим Международной организацией по стандартизации (ISO – *International Standard Organisation*) разработана концепция глобальной сети связи общего пользования. Под сетью понимается множество взаимодействующих друг с другом систем, осуществляющих хранение, обработку и передачу разнообразной информации.

Глобальная сеть объединяет телефонные сети общего пользования, сети передачи данных, системы мобильной и спутниковой связи. Ею охватывается все многообразие пользовательских терминалов (проводные и бесшнуровые телефонные аппараты, персональные компьютеры, сотовые и спутниковые телефоны, аппараты факсимильной связи и пр.), линий связи (кабельных, оптических, беспроводных и т.д.), узлов коммутации (учрежденческих, районных, междугородних), систем управления, контроля и технического обслуживания. Для того, чтобы составить адекватное представление о наиболее общих чертах и архитектуре столь сложного и разветвленного образования, разумно вписать его в рамки некоторой абстрактной модели, классифицирующей элементы сети по их функциональному назначению и механизмам взаимодействия.

4.1 Понятие о модели взаимодействия открытых систем

Эта эталонная модель (OSI – *Open Systems Interconnection*), разработанная ISO, обычно обозначается как модель ISO/OSI и является базовой в том смысле, что для описания некоторых систем достаточна лишь часть ее компонентов. Несмотря на существование конкурентоспособных альтернатив, модель OSI получила повсеместное распространение и широко используется как концептуальная основа в информационной системотехнике.

Под *открытой* системой понимается система, удовлетворяющая требованиям международных организаций или требованиям производителя, контролирующего большую часть рынка сбыта. Открытость означает, что система не является чьей-то собственностью, т.е. любой производитель вправе создавать и выставлять на продажу собственный продукт (аппаратуру или программное обеспечение), предназначенный для практического доступа к услугам системы. Между производителями возникает конкуренция, что способствует снижению цены оборудования. Пользователи при этом не зависят от единственного производителя и могут менять поставщиков аппаратуры.

Диаграмма рис. 4.1 иллюстрирует схему взаимодействия элементов и уровней базовой модели, представленной в общем виде. Модель некоторых систем содержит лишь часть этих элементов. Так, системы с коммутацией каналов описываются одноуровневой моделью. Расшифруем используемые термины.

Прикладной процесс выполняет обработку данных для нужд пользователя либо с помощью некоторых аппаратно-программных средств, либо непосредственно в мозгу человека.

Под *интерфейсом* (стыком) понимается определенная стандартами граница между взаимодействующими объектами. Для систем мобильной связи главенствующее значение имеет радиointерфейс, охватывающий средства соединения между мобильной и базовой станциями.

Физические средства соединения - это совокупность аппаратуры и среды распространения, обеспечивающих передачу сигналов между системами.

Абонент через интерфейс пользователя дает задание прикладному процессу и получает результат его выполнения.

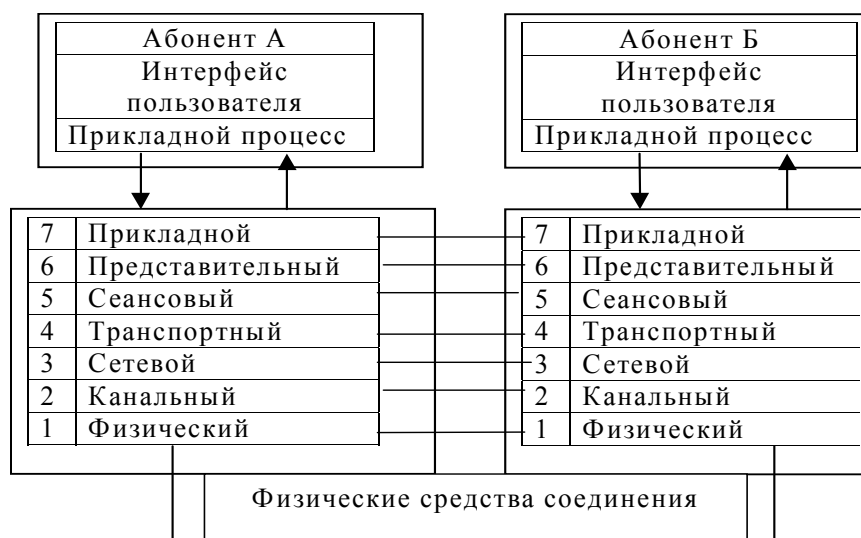


Рис.4.1. Модель взаимодействия открытых систем

Областью взаимодействия называется иерархическая группа функциональных блоков, предназначенных для обмена данными между системами. В базовой модели эти функции поделены на семь расположенных друг над другом слоев, именуемых *уровнями*, как показано на рис. 4.1.

Такое деление позволяет разбить сложную систему на ряд более простых (осуществить декомпозицию), образовать стандартные модули и интерфейсы между уровнями. Опишем кратко основные функции, закрепленные за каждым уровнем.

Верхний, седьмой прикладной уровень обеспечивает различные формы взаимодействия прикладных процессов, т. е. является прикладным интерфейсом. На этом уровне выполняется управление заданиями, идентификация пользователей по паролям, адресам, электронным подписям.

На представительном уровне осуществляется кодирование сообщений, поступающих с седьмого уровня, в нужную форму. Здесь же при необходимости осуществляется сжатие данных и шифрование.

Сеансовый уровень определяет процедуру проведения сеансов между прикладными процессами, т. е. обеспечивает установку, поддержание и разрыв соединения.

Транспортный уровень отвечает за точность доставки пакетов (блоков данных по 500 – 2000 бит), по виртуальным каналам, проложенным между портами.

Сетевой уровень создает виртуальные каналы – пути, по которым данные передаются от одного порта к другому.

На канальном уровне производится формирование и передача кадров (блоков данных), обнаружение и исправление ошибок.

Физический уровень создает интерфейсы для подключения систем к физическим средствам соединения (электрическим и оптическим кабелям, радиоэфиру вместе с аппаратурой передачи данных). Он отвечает за передачу последовательности бит, оповещает о появлении неисправностей и отказов (столкновении кадров, обрывов, отключении питания).

Хотя все сообщения передаются через первый – физический уровень, в модели OSI определены также *логические* каналы, понимаемые как пути, по которым передаются сигналы или данные между одноименными уровнями. Так, путь, связывающий четвертые уровни, называется транспортным каналом, путь между третьими – сетевым, вторыми – информационным и, наконец, собственно физическим каналом, называется путь, соединяющий первые уровни. Особое место занимает логический канал *трафика* (пользовательский канал), обеспечивающий обмен данными между абонентами.

При технической реализации систем логические каналы проектируются на физические, т. е. указывается, какой физический канал в данный момент времени выделяется для организации конкретного логического канала.

Эталонная модель обладает свойствами *прозрачности* и *доступности* сервиса. Прозрачность означает, что данные, переданные с какого-то уровня на нижние, на приемной стороне возвращаются на этот же уровень без изменений. Создается видимость, что для рассматриваемых одноименных уровней нижних не существует, они прозрачны. Ярким примером прозрачного объекта может служить телефонный тракт. Большинство людей, пользующихся телефоном, даже не имеют представления о преобразованиях, которым подвергается их речь в телефонном канале.

Под доступностью сервиса понимается возможность использования всеми верхними уровнями услуг (сжатие данных, защита от помех, от несанкционированного использования информации), выполняемых на нижних уровнях.

На рис. 4.1 линиями, соединяющими одноименные уровни, обозначены *протоколы* – правила взаимодействия двух (или более) объектов. Названия протоколов совпадают с названиями соответствующих уровней, например, существуют сеансовые, транспортные, сетевые протоколы.

Протоколы должны удовлетворять определенным требованиям. Прежде всего, протокол – это последовательность шагов, выполняемых взаимодействующими объектами добровольно, в порядке строгой очередности для достижения поставленной цели. Далее, протокол должен содержать описание действий его участников при любых возникающих ситуациях (должны быть исключены тупиковые ситуации). Правила игры в шахматы, в шашки являются примерами протоколов.

Следует упомянуть о существовании криптографических протоколов, обеспечивающих достижение цели в условиях возможного мошенничества, как со стороны законных пользователей, так и внешнего нарушителя. Примером такого протокола может служить аутентификация абонента.

Для описания систем мобильной связи базовая семиуровневая модель трансформируется путем деления на два укрупненных слоя. В стандартах GSM и TETRA верхние уровни с четвертого по седьмой образуют один слой, нижние (с первого по третий) – другой. При передаче речи функции седьмого и шестого уровней выполняет человек. В его мозгу зарождается сообщение, он выбирает язык общения. Задачи, решаемые на пятом и четвертом уровнях, важны для соединительных линий, объединяющих отдельные системы связи в единую сеть. Поэтому указанные системы описываются трехуровневой моделью, включающей лишь сетевой, каналный и физический уровни.

4.2 Структура сетей мобильной связи

Основные, наиболее общие элементы сети связи показаны на рис. 4.2. Они содержат мобильные станции (МС), базовые станции (БС), центры управления и коммутации (ЦК), радиоинтерфейсы (РИ). МС используются подвижными абонентами. БС – стационарные приемопередающие станции, обеспечивающие радиосвязь с МС в определенной зоне (ячейке, соте, секторе). В ЦК выбирается направление передачи данных (коммутация), производится управление БС и осуществляется доступ к внешним системам.

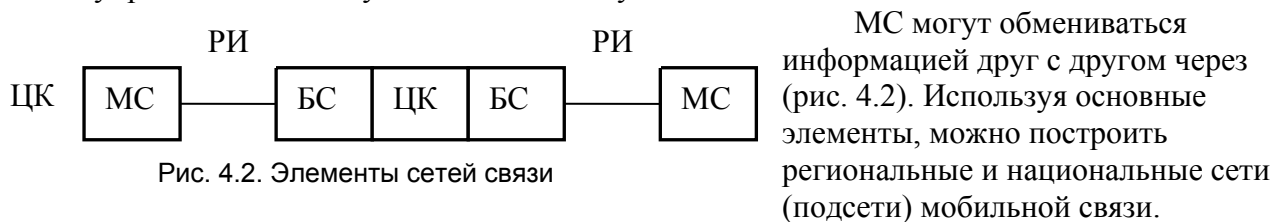


Рис. 4.2. Элементы сетей связи

Отдельные *сети* и *подсети* объединяются в единую структуру с помощью межсетевых интерфейсов (СИ) – кабельных или радиолоний. Контроллеры базовых станций (КБС) вы-

полняют функции коммутации, управления БС, обеспечивают доступ к внешним сетям (телефонным, передачи данных). Они могут подразделяться на главные (с расширенными возможностями) и подчиненные (с ограниченными функциями). Подсеть может содержать собственную станцию управления (СУ). Таким образом, функции управления и коммутации распределяются по всей сети, что обеспечивает быструю передачу вызовов и сохранение работоспособности сети при неисправностях отдельного оборудования.

4.3 Классификация и эволюция систем мобильной радиосвязи

4.3.1 Основные типы систем мобильной связи

Приведем в кратком изложении следующий основной набор классификационных признаков систем мобильной связи (СМР):

- способ управления системой, иначе способ объединения абонентов – *централизованный* (координированный) или *автономный* (некоординированный). При централизованном объединении связь между абонентами производится через центральные (или базовые) станции. В противном случае связь между пользователями устанавливается непосредственно, без участия базовых станций;
- зона обслуживания – *радиальная* (в пределах радиуса действия радиостанции), *линейная* (для линейно протяженных зон), *территориальная* (для определенных конфигураций территории);
- направленность связи – *односторонняя* или *двусторонняя* связь между абонентом и базовой станцией;
- вид работы системы – *симплекс* (поочередная передача от абонента к базовой станции и обратно) или *дуплекс* - одновременная передача и прием в каждом из двух названных направлений;
- метод разделения каналов в системе радиосвязи, или метод множественного доступа – *частотный*, *временной* или *кодовый*;
- способ использования частотного ресурса, выделенного системе связи – жесткое закрепление каналов за абонентами, возможность доступа абонентов к общему частотному ресурсу (транкинговые системы), повторное использование частот за счет пространственного разнесения передатчиков (сотовые системы);
- категория обслуживаемых системой связи абонентов – профессиональные (служебные, корпоративные) абоненты, частные лица;
- вид передаваемой информации – речь, кодированное сообщение и др.

В зависимости от назначения системы, объема предоставляемых услуг и размеров зоны обслуживания можно выделить следующие четыре типа СМР:

- транкинговые системы связи (ТСС);
- системы персонального радиовызова (СПРВ);
- системы персональной спутниковой связи (СПСС);
- сотовые системы мобильной связи (ССМС).

По способу организации множественного доступа, т. е. технологии распределения между отдельными каналами связи частотно-временного ресурса, выделяют СМР на основе одной из трех конкурирующих технологий:

- множественный доступ с частотным разделением каналов (МДЧР, англоязычная аббревиатура FDMA – *frequency division multiple access*);
- множественный доступ с временным разделением каналов (МДВР или TDMA – *time division multiple access*);
- множественный доступ с кодовым разделением каналов (МДКР или CDMA – *code*

division multiple access);

4.3.2 Транкинговые системы связи

Термин "*транкинговая*" (или *транковая*) связь происходит от английского слова *trunk* (ствол) и отражает то обстоятельство, что "ствол связи" содержит несколько каналов, причем жесткое закрепление каналов за абонентами отсутствует. В литературе можно найти различные определения транкинговых систем, общим для которых является именно предоставление в распоряжение абонента одного из свободных на данный момент каналов.

Типичной сферой применения транкинговых систем являются государственные, ведомственные, корпоративные организации и институты, такие как служба скорой помощи, пожарная служба, охрана правопорядка, органы безопасности, различные коммерческие структуры и др. По большей части транкинговые системы используются как средства оперативной связи с жестко лимитированным и постоянно контролируемым контингентом абонентов в пределах ограниченной территориальной зоны.

Транкинговые системы связи могут строиться как системы с однозоновой или многозоновой структурой. Принимая во внимание специфический характер ТСС, т.е. ограниченность числа пользователей системы, переход от однозоновой к многозоновой структуре объясняется в первую очередь расширением географической зоны действия системы, а не стремлением к повышению числа абонентов (абонентской емкости) системы. При пересечении границ радиопокрытия ТСС отслеживают перемещение абонентов, обеспечивают их регистрацию, и назначение им нового частотного канала. Однако, как правило, подобный переход происходит с прерыванием связи, для восстановления которой абонентам необходимо произвести повторный вызов.

Транкинговые системы могут использовать как симплексные, так и дуплексные каналы радиосвязи, однако с целью упрощения и удешевления в них нередко применяется полудуплексный режим работы, при котором один и тот же канал поочередно используется для связи от центра управления (базовой станции) к абоненту и в обратном направлении.

Реализация принципа равного доступа к каналу связи может быть осуществлена децентрализованно либо при централизованном управлении. В первом случае функция нахождения свободного канала возлагается на абонентскую станцию, которая проводит последовательный поиск незанятого частотного канала во всем выделенном системе диапазоне. Во втором случае анализ занятости каналов связи осуществляет базовая станция либо непосредственно центр коммутации мобильной связи. Как правило, установление связи при последовательном сканировании частотного диапазона занимает достаточно большой интервал времени. Для обеспечения оперативности управления в современных ТСС предусматривается существование специального канала, посредством которого производится управление транкинговой системой, в том числе выполнение процедур установления и прекращения связи.

По способу организации канала управления различают ТСС с выделенным и распределенным каналом управления. В первом случае, как следует из названия, выделенный канал используется исключительно для управления работой системы. Во втором – в процессе сеанса связи сигналы управления передаются одновременно с речевым сигналом.

С учетом сказанного транкинговая система связи может быть представлена обобщенной структурной схемой (см. рис.4.3), где использованы следующие обозначения:

МС – мобильная станция (мобильный абонент);

БС – базовая станция (центр управления);

УОР – устройство объединения радиосигналов;

Р – ретрансляторы;

ЦКПС – центр коммутации мобильной связи;

ТФОП – телефонная сеть общего пользования;

ДПУ – диспетчерский пункт управления.

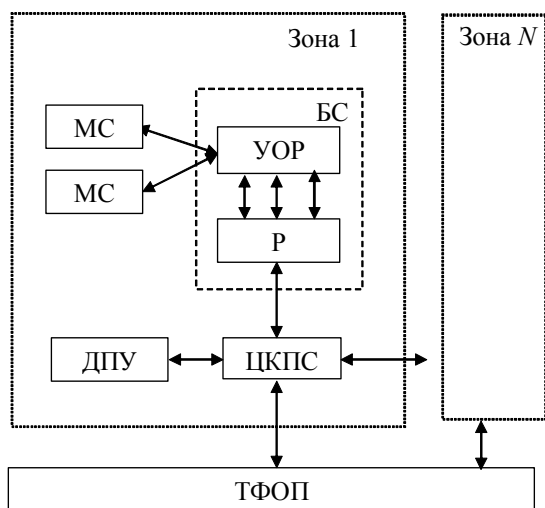


Рис. 4.3. Структурная схема ТСС

Следует отметить, что для ТСС наиболее характерно разделение каналов связи по частоте с индивидуальными ретрансляторами на разных частотах. Возможен и вариант ТСС с использованием широкополосных ретрансляторов, обслуживающих сразу все каналы. Назначение остальных блоков структурной схемы является очевидным и не требует дополнительных комментариев.

Стандарт MPT 1327, разработанный министерством почт и телекоммуникаций Великобритании (*Ministry of Post and Telecommunication* (MPT)), определяет в основном протокол передачи информации управления и контроля состояния аппаратуры (иначе - информации сиг-

нализации) для транкинговых систем наземной мобильной радиосвязи, причем информационные сообщения передаются по аналоговому радиоканалу. На его основе разработаны радиоинтерфейс МС (абонента), определяемый протоколом MPT 1343, и радиоинтерфейс БС – MPT 1347. Стандартами предусматривается передача информации со скоростью 1,2 кбит/с по каждому из 500 каналов связи в диапазоне частот 201,2125...207,4875 МГц (MPT 1347) и 193,2125...199,4875 МГц (MPT 1343), причем каждый дуплексный канал занимает две полосы шириной 12,5 кГц с разносом каналов приема и передачи в 8 МГц.

Фирмой Ericsson разработана система транкинговой радиосвязи, получившая название EDACS (*Enhanced Digital Access Communications System* – Усовершенствованная система связи с цифровым доступом). Системы EDACS выпускаются в различных модификациях, причем различают "системы EDACS", "сети EDACS" и "расширенные сети EDACS". Системы EDACS, объединенные между собой посредством контроллеров узлов связи и диспетчерских пунктов управления, образуют сети EDACS, которые, в свою очередь, с помощью некоторых интегрированных узлов связи могут объединяться в расширенную сеть для покрытия значительных территорий.

В системе EDACS применяются два типа радиоканалов – рабочий канал и канал управления. Канал управления служит для обмена цифровой информацией сигнализации между мобильными станциями и устройствами управления работой всей системы. Рабочие каналы используются для обмена собственно информацией (разговорной или данными) между мобильными станциями. Системы и сети EDACS рассчитаны на использование как аналоговых, так и цифровых станций, обеспечивающих передачу речевых сигналов в цифровой форме. Стандартная скорость передачи данных составляет 9,6 кбит/с по каждому из 20 каналов системы EDACS в диапазонах частот 30...300 МГц, 800 МГц или 900 МГц с разносом каналов связи 25, 30 и 12,5 кГц.

Общие тенденции, связанные с унификацией и интеграцией СМР идентичного назначения, привели к разработке под эгидой ETSI (*European Telecommunications Standards Institute* – Европейский институт стандартов связи) общеевропейского стандарта TETRA (*Trans-European Trunked Radio* – Общеевропейская система транкинговой связи), изменившего свое название с апреля 1997 г. на *TErrestrial Trunked Radio* (Сухопутная система транкинговой связи) ввиду своего широкого распространения. ТСС на основе стандарта TETRA представляют собой новое поколение систем этого типа, следующее за аналоговым. В отличие от предыдущих в стандарте TETRA осуществлен полный переход к цифровому представлению передаваемой информации и использовано не частотное, а временное разделение каналов. В результате этих и ряда других мер скорость передачи в системе достигла 36 кбит/с. Для сис-

темы TETRA выделены два дуплексных участка спектра в полосе частот 380...400 МГц при разносе радиоканалов для приема и передачи в 10 МГц и разносе соседних каналов в 25 кГц.

4.3.3 Системы персонального радиовызова

Системы персонального радиовызова (СПРВ), или *пейджинговые* системы (от английского термина *paging* – вызов) представляют собой системы мобильной радиосвязи, обеспечивающие одностороннюю передачу коротких сообщений из центра системы (терминала персонального радиовызова или пейджингового терминала) на миниатюрные абонентские приемники (пейджеры), причем передаваемые сообщения могут быть четырех типов – тональные, цифровые, буквенно-цифровые и речевые. В последнее время стали появляться публикации о разработке пейджеров для двусторонней связи, способных осуществлять передачу подтверждений о приеме сообщений [11].

По своему назначению СПРВ можно подразделить на частные (ведомственные) и общего пользования. Частные СПРВ предназначены для передачи сообщений на ограниченной территории в интересах отдельных групп пользователей, причем, как правило, передача осуществляется с диспетчерского терминала радиовызова без взаимодействия с ТФОП. Под системами общего пользования понимают системы мобильной радиосвязи, обеспечивающие передачу сообщений ограниченного объема от пользователей ТФОП (или с персонального компьютера) через пейджинговый терминал на абонентские приемники СПРВ. Потенциальными пользователями СПРВ в основном являются различные экстренные службы (скорая помощь, пожарная и аварийно-спасательная службы, полиция и пр.), а также лица, занятые в сфере оказания различного рода услуг (консультационных, бытового технического обслуживания и др.).

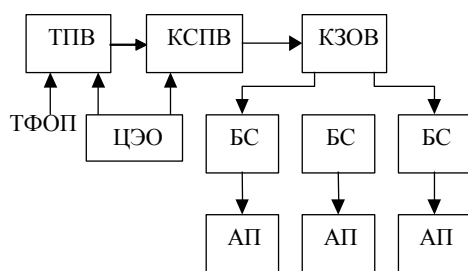
Основной отличительной чертой СПРВ служит ее работа вне реального времени, т.е. сообщения передаются не в момент его выдачи отправителем, а в порядке очереди, устанавливаемой центром управления, хотя на практике величина задержки между моментом отправления и получения не превышает нескольких минут. В сочетании с краткостью сообщений и односторонней направленностью СПРВ оказываются технически очень простыми, а значит и экономически выгодными потребителю, поэтому системы пейджинговой связи нередко называют "мобильной связью для бедных".

В самой общей версии структурная схема СПРВ может быть представлена рис.4.4, где используются сокращения:

ТПВ – терминал персонального вызова (*paging terminal* – PT),

КСРВ – контроллер сети персонального вызова (*paging network controller* – PNC);

ЦЭиО – центр эксплуатации и обслуживания (*operations and maintenance center* – ОМС);



КЗОВ – контроллер зоны обслуживания вызовов (*paging area controller* – PAC);

БС – базовая станция (*base station* – BS);

АП – абонентский приемник (*pager* – P).

Терминал персонального вызова принимает и анализирует входящие данные вызовов из ТФОП (или от других источников). Контроллер сети персонального вызова во взаимодействии с ЦЭиО обеспечивает все функции управления системой, в том числе такие, как определение типа вызова абонента – индивидуальный, общий или групповой. Контроллер зоны обслуживания

Рис. 4.4. Структурная схема СПРВ

осуществляет распределение данных к передатчикам базовых станций, а также может выполнять некоторые статистические вычисления. Базовые станции контролируют и передают сигналы радиовызова на соответствующие абонентские приемники. Применение нескольких БС в пределах зоны действия СПРВ позволяет обеспечить более надежную связь с абонентами.

Примером дальнейшего развития СПРВ может служить система протоколов FLEX (включая ReFLEX и InFLEXion), разработанная фирмой Motorola. Достоинством последних является повышенная скорость передачи сообщений, возможность двустороннего пейджинга, а также передача голосовых сообщений.

На ранних этапах развития массовой мобильной связи системы персонального радиовызова получили достаточно широкое распространение, причем рост количества обслуживаемых абонентов доходил до 20% ежегодно. Так, если в 1990 г. в мире насчитывалось 20 млн. пользователей, то в 1997 г. их число превысило 100 млн.. В России на начало 21-го столетия пейджинговая абонентская сеть исчислялась многими сотнями тысяч. Вместе с тем для стран с высоким уровнем проникновения мобильного телекоммуникационного сервиса (ЕС, Скандинавия, США и др.) характерно постепенное свертывание сетей персонального радиовызова и замещение их услуг более прогрессивными и постоянно дешевеющими системами сотового телефона.

4.3.4 Сотовые системы мобильной связи

Рассмотренные в разделе 2.2 ТСС в принципе не ориентированы на коммерческого потребителя, имея первоочередной целью удовлетворение ведомственных и корпоративных нужд. Они, как правило, характеризуются небольшой абонентской емкостью, фиксированной и весьма скромной номенклатурой услуг, невысокой скоростью передачи данных и пр. СПРВ тем более не отвечают потребительским стандартам 21-го века и имеют статус переходного средства одностороннего оповещения, постепенно вытесняемого более прогрессивными технологиями мобильной связи по мере расширения и удешевления последних. Что же касается спутниковых персональных систем связи, то основным препятствием на пути их массового внедрения следует считать экономический фактор, т. е. высокую стоимость предоставляемых услуг.

В свете сказанного становятся понятными причины, по которым ведущие позиции в коммерческой мобильной связи принадлежат наземным системам, основанным на возможности многократного использования выделенного ресурса при соответствующем пространственном разнесении приемопередатчиков базовых станций. Исторически первыми системами, многократно эксплуатирующими выделенный ресурс, явились системы с повторным использованием частотных каналов. Идея повторного применения частот заключается в том, что в смежных областях радиопокрытия СМР используются разные полосы разрешенного частотного диапазона, тогда как в зонах, достаточно удаленных друг от друга, допускается передача в одних и тех же частотных каналах. Возможность подобного частотно-территориального планирования объясняется быстрым пространственным затуханием радиоволн дециметрового диапазона, применяемых в СМР этого типа. Участок территории радиопокрытия, на котором осуществляется связь в фиксированной полосе частот, схематически изображается в виде правильного шестиугольника и по сходству с пчелиными сотами получил название *соты*. В результате СМР с пространственным разнесением частот получили наименование *сотовых систем мобильной связи* (ССМС). Группу сот, в пределах которой отсутствует повторное использование частотных полос, называют *кластером*. Сотовая топология позволяет многократно увеличить абонентскую емкость системы по сравнению с системами радиальной структуры и охватить сколь угодно большую зону обслуживания без ухудшения качества связи и расширения выделенного частотного диапазона. Вместе с тем, использование сотового принципа построения предполагает и ряд усложнений, касающихся определения текущего местоположения мобильного абонента и обеспечения непрерывности связи при перемещении его из одной соты в другую. Соответствующая процедура получила название *эстафетной передачи* (в английской транскрипции *handoff* или *handover*).

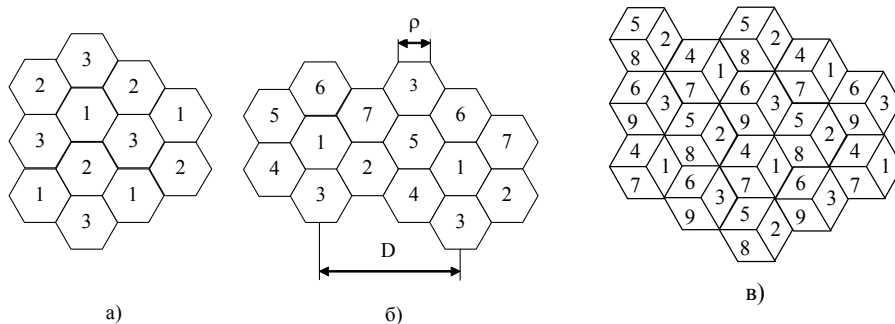


Рис. 4.5. Варианты структур кластера ССМС

Высокая спектральная эффективность ССМС достигается ценой максимально частого повторного использования одних и тех же частотных полос, и с этой точки зрения наиболее предпочтительным был бы трехсотовый (или трехэлементный) кластер, изображенный на рис.2.3 а, где одинаковыми цифрами обозначены соты с совпадающими наборами частотных каналов. Кроме того, каждой из сот кластера данного типа отводится частотная полоса, равная трети полного частотного диапазона, а значит, и треть общего числа каналов связи в системе, что обеспечивает значительную абонентскую емкость соты. Вместе с тем, частое повторение зон с одинаковыми полосами частот характеризуется заметным уровнем соканальных помех, т.е. помех от станций системы, работающих в той же полосе частот, но расположенных в несмежных сотах. Для уменьшения влияния соканальных помех более выгодны кластеры с большим числом элементов, например 7-элементные, изображенные на рис.4.5 б. Можно показать, что расстояние D между центрами ячеек, в которых используются одинаковые полосы частот, и число n_c элементов в кластере связаны соотношением

$$D = \sqrt{3n_n} \rho,$$

где ρ - радиус ячейки, т.е. радиус окружности, описанной вокруг правильного шестиугольника. Параметр ξ , определяемый соотношением

$$\xi = \frac{D}{\rho} = \sqrt{3n_n},$$

называют *коэффициентом уменьшения соканальных помех* или *коэффициентом соканального повторения*. Для величины $\eta = 1/n_c$ употребляют наименование *коэффициент эффективности повторного использования частот* или *коэффициент повторного использования частот*. Увеличение числа элементов в кластере, благоприятно сказывающееся на уровне соканальных помех, приводит к пропорциональному уменьшению полосы частот, которая может быть использована в каждой соте, а значит к снижению абонентской емкости соты.

Рассмотренные структуры кластеров предполагают использование на базовых станциях антенн с круговой диаграммой направленности, осуществляющих передачу сигнала по всем направлениям с одинаковой мощностью. Эффективным способом снижения соканальных помех является применение направленных (в горизонтальной плоскости) антенн с шириной диаграммы направленности 120° или 60° , в результате чего шестиугольная ячейка разбивается на 3 или 6 секторов, т.е. производится *секторизация сот*. В секторе сигнал излучается антенной только в одну сторону, а уровень излучения в противоположном направлении сокращается до минимума. Таким образом, секторизация сот позволяет чаще использовать одинаковые полосы частот в кластерах без изменения их структуры либо в рамках прежней схемы повторения частот заметно снизить уровень соканальных помех. Если кластер состоит из n_c сот, каждая из которых содержит m_c секторов, то говорят, что размерность кластера $(n_n, n_n \times m_n)$. Типичными размерностями кластеров, широко применяемых на практике, являются (3,9), (4,12), (7,21). На рис.4.5 в приведен возможный вариант распределения наборов частотных каналов в секторизованном кластере размерности (3,9). Отметим, что разработка топологии ССМС является своеобразной и достаточно сложной задачей.

4.3.5 Эволюция систем и стандартов сотовой связи

История развития ССМС насчитывает немногим более 25 лет, однако, этот короткий период был отмечен рядом поворотных моментов и существенной эволюцией воззрений на роль и философию рассматриваемых систем. Можно говорить о трех поколениях ССМС, различия между которыми – с известной долей условности – устанавливаются следующими критериями.

Все ССМС или стандарты первого поколения являются аналоговыми. В их числе:

- AMPS (*Advanced Mobile Phone Service*) – Усовершенствованная мобильная телефонная служба. Диапазон рабочих частот – 869...894 МГц для БС и 824...849 МГц для МС, ширина полосы канала связи – 30 кГц. Начало коммерческого применения – 1983 г. Широко использовалась в США, Канаде, Центральной и Южной Америке, Австралии.

- TACS (*Total Access Communications System*) – Общедоступная система связи. Частотный диапазон: 935...950 МГц для БС, 890...905 МГц для МС, ширина полосы канала связи – 25 кГц. Начало коммерческого применения – 1985 г. Наибольшее распространение стандарт TACS получил в европейских странах – Англия, Италия, Испания, Австрия и др.

- NMT (*Nordic Mobile Telephone System*) – Скандинавская система мобильной телефонной связи. Существует в двух основных вариантах NMT 450 и NMT 900, differing только диапазоном используемых частот: NMT 450 – 463...467,5 МГц для БС и 453...457,5 для МС; NMT 900 – 935...960 МГц для БС и 890...915 МГц для МС. Ширина полосы канала – 25 кГц. Начало коммерческого использования – 1981 г. (NMT 450) и 1986 г. (NMT 900). Помимо скандинавских стран эти стандарты широко использовались во многих странах Западной и Восточной Европы и ряде других регионов мира.

Во всех перечисленных аналоговых стандартах применяется метод множественного доступа с частотным разделением каналов, для передачи речевой информации используется частотная модуляция, а для передачи информации сигнализации – частотная манипуляция. Системам первого поколения присущ ряд недостатков, основными из которых являются относительно низкая абонентская емкость, несовместимость различных стандартов, отсутствие засекречивания передаваемых сообщений, невозможность взаимодействия с цифровыми системами с интеграцией служб (ISDN) и пакетной передачей данных (PDN).

В значительной степени указанных недостатков лишены цифровые ССМС второго поколения, среди которых наибольшее распространение получили следующие:

- D-AMPS (*Digital-AMPS* – цифровая AMPS), или IS-54 (IS – сокращение от *Interim Standard*, т.е. промежуточный стандарт) представляет двухрежимную аналого-цифровую систему, совмещающую работу в аналоговом и цифровом режимах в том же диапазоне, что и AMPS. Начало практического использования относится к 1992 г. Усовершенствованная версия данного стандарта IS-136, отличие которой от IS-54 заключается в наличии полностью цифровых каналов управления, начала применяться с 1996 г. Версия IS-136 используется в диапазонах 800 и 1900 МГц.

- GSM (*Global System for Mobile Communications*) – Глобальная система мобильной связи. Данным стандартом предусматривается работа в диапазоне 935...960 МГц для БС и 890...915 МГц для МС при ширине полосы канала связи 200 кГц. Практическое применение общеевропейского стандарта GSM 900 началось в 1991 г. Совершенствование данного стандарта привело к освоению нового частотного диапазона 1800 МГц, в котором благодаря более широкой рабочей полосе частот в сочетании с меньшими размерами сот удается строить сотовые сети значительно большей емкости. Первоначально данная версия именовалась *Personal Communication Network (PCN)* – Сеть персональной связи, затем *Digital Cellular System (DCS)* – Цифровая система сотовой связи, а спустя три года после начала эксплуатации (1993 г.) была переименована в GSM 1800. Диапазон работы БС – 1710...1785 МГц, МС – 1805...1880 МГц при ширине полосы канала связи 200 кГц. Стандарт GSM нашел применение и в США, однако, из-за того, что диапазон 1800 МГц занят системой D-AMPS в версии IS-136, ему была выделена полоса частот в диапазоне 1900 МГц. Соответствующая версия стандарта GSM получила наименование "американский" GSM или IS-661. На начало 1999 г.

стандарт GSM в различных версиях использовался в 129 странах мира, а объем абонентской базы достиг величины 137 млн. чел., что составляет 45% от общего числа пользователей ССМС.

- PDC (*Personal Digital Cellular*) – Персональная цифровая сотовая связь. Цифровая ССМС, разработанная в Японии в 1993 г. и первоначально называвшаяся JDC (*Japan Digital Cellular* – Японская цифровая сотовая связь), по своим техническим характеристикам подобна D-AMPS и отличается от последней возможностью работы в нескольких диапазонах частот – 800, 1400 и 1500 МГц. Хотя стандарт PDC используется только в Японии, его абонентская база достигает 39,5 млн. пользователей или 13% от общемирового числа абонентов ССМС.

- IS-95 (*Interim Standard-95*), альтернативное наименование cdmaOne. Все перечисленные ранее цифровые ССМС второго поколения используют метод множественного доступа с частотно-временным разделением каналов связи. Критическими явились 1992-93 г.г., когда в США был разработан первый стандарт ССМС на основе метода множественного доступа с кодовым разделением (CDMA), получивший название IS-95. Диапазон рабочих частот –824...848 МГц для МС и 869...894 МГц – для БС при ширине спектра излучаемых сигналов 1,25 МГц. Практическое применение указанного стандарта началось в 1995-96 г.г. в Гонконге, США и Южной Корее, причем в США используется версия этого стандарта для диапазона 1900 МГц.

Что касается ССМС третьего поколения, то процесс формирования их облика еще продолжается, изобилуя коллизиями, конфликтами интересов и компромиссами. Принципиальные критерии, отличающие ССМС третьего поколения, а также некоторые детали конкретных спецификаций, имеющих фактический статус стандартов и весьма близких к практическому воплощению, рассмотрены в разделе 4.7.

4.3.6 Описание ССМС на примере GSM

Рассмотрим основные элементы ССМС и характер их взаимодействия. Хотя сотовые системы разных стандартов имеют значительные отличия в деталях, интегральное их описание с помощью некоторой обобщенной модели возможно и полезно. В качестве подобного обобщения выберем схему ССМС, приближенную к стандарту GSM, как наиболее отчетливо структурированную. Функциональное построение ССМС иллюстрируется схемой на рис. 4.6. В приведенной схеме условно можно выделить четыре основных компонента – центр управления и обслуживания (ЦУО) (иначе *operations and maintenance center* (ОМС)) и три подсистемы:

- подсистема мобильных станций ПМС (*mobile station subsystem* MSS);
- подсистема базовых станций ПБС (*base station subsystem* BSS);
- подсистема коммутации ПК (*switching subsystem* SSS),

функциональное сопряжение которых описывается рядом интерфейсов.

ЦУО является центральным элементом ССМС, который обеспечивает управление другими компонентами системы, а также контроль качества функционирования. Подсистема мобильных станций объединяет оборудование, обеспечивающее доступ абонентов в систему. Главным звеном в архитектуре ССМС является подсистема коммутации, которая включает в себя центр коммутации мобильной связи ЦКМС (*mobile switching center* MSC), визитный (гостевой) регистр местоположения ВРМ (*visited location register* VLR), домашний регистр местоположения ДРМ (*home location register* HLR), центр аутентификации ЦА (*authentication center* AUC) и регистр идентификации оборудования РИО (*equipment identity register* EIR). В подсистему базовых станций входят базовые приемо-передающие станции БС (*base transceiver station* BTS) и контроллеры базовых станций КБС (*base station controller* BSC).

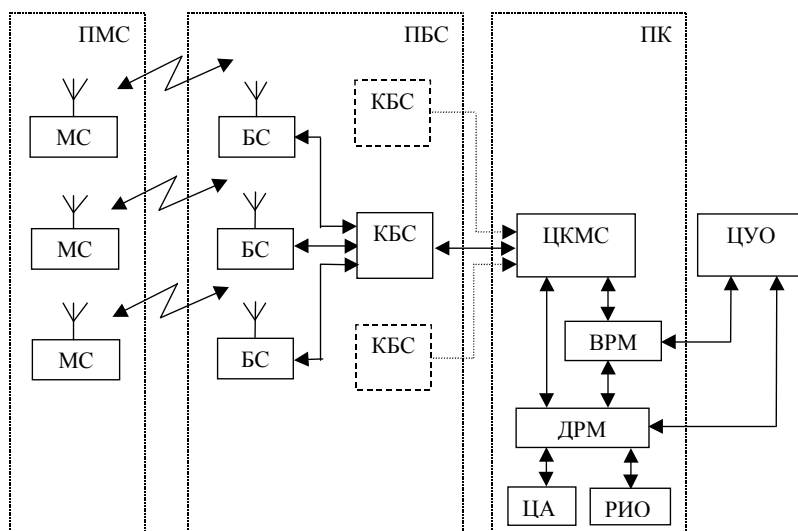


Рис. 4.6. Обобщенная структурная схема сотовой системы мобильной связи

Центр коммутации мобильной связи обслуживает группу сот и обеспечивает все виды соединений, в которых нуждается в процессе работы мобильная станция, т.е. коммутацию мобильных абонентов друг с другом, с абонентом ТФОП и др. На ЦКМС возлагаются также функции коммутации каналов, к которым относятся "передача обслуживания" (или "эстафетная передача") и переключение каналов в соте при появлении сильных помех и неисправностей, если только это не является обязанностью КБС. Помимо коммутационных задач ЦКМС управляет процедурами слежения за мобильными станциями с помощью домашнего и визитного регистров местоположения для обеспечения доставки вызова, а также процедурами аутентификации и идентификации абонентов с помощью ЦА и РИО.

Когда передача вызовов осуществляется между двумя сетями, управляемыми разными КБС (BSC), то первичное управление осуществляется в ЦКМС (MSC). В стандарте GSM также предусмотрены процедуры передачи вызова между сетями (контроллерами), относящимися к разным ЦКМС (MSC). Центр коммутации осуществляет постоянное слежение за подвижными станциями, используя регистры положения ДРМ (HLR) и перемещения ВРМ (VLR). В HLR хранится та часть информации о местоположении какой-либо подвижной станции, которая позволяет центру коммутации доставить вызов станции. Регистр HLR содержит международный идентификационный номер подвижного абонента (IMSI). Он используется для опознавания подвижной станции в центре аутентификации ЦА (AUC).

Регистры ДРМ (HLR) и ВРМ (VLR) по своей сути представляют собой базы данных. Первый содержит сведения о постоянно приписанных к данному ЦКМС абонентах и о видах услуг, которые им могут быть оказаны, второй содержит информацию об абонентах, временно находящихся в зоне обслуживания данного ЦКМС (MSC). Центр аутентификации обеспечивает возможность проведения процедуры аутентификации абонентов и шифрования передаваемых сообщений, РИО содержит сведения об эксплуатируемых мобильных станциях на предмет их исправности и санкционированного использования.

К данным, содержащимся в HLR, имеют дистанционный доступ все MSC и VLR сети и, если в сети имеются несколько HLR, в базе данных содержится только одна запись об абоненте, поэтому каждый HLR представляет собой определенную часть общей базы данных сети об абонентах. Доступ к базе данных об абонентах осуществляется по номеру IMSI или MSISDN (номеру подвижного абонента в сети ISDN). К базе данных могут получить доступ MSC или VLR, относящиеся к другим сетям, в рамках обеспечения межсетевого роуминга абонентов.

Второе основное устройство, обеспечивающее контроль за передвижением подвижной станции из зоны в зону, - регистр перемещения VLR. С его помощью достигается функционирование подвижной станции за пределами зоны, контролируемой HLR. Когда в процессе

перемещения подвижная станция переходит из зоны действия одного контроллера базовой станции BSC, объединяющего группу базовых станций, в зону действия другого BSC, она регистрируется новым BSC, и в VLR заносится информация о номере области связи, которая обеспечит доставку вызовов подвижной станции. Для сохранности данных, находящихся в HLR и VLR, в случае сбоев предусмотрена защита устройств памяти этих регистров.

В сети подвижной связи GSM соты группируются в географические зоны (LA), которым присваивается свой идентификационный номер (LAC). Каждый VLR содержит данные об абонентах в нескольких LA. Когда подвижный абонент перемещается из одной LA в другую, данные о его местоположении автоматически обновляются в VLR. Если старая и новая LA находятся под управлением различных VLR, то данные на старом VLR стираются после их копирования в новый VLR. Текущий адрес VLR абонента, содержащийся в HLR, также обновляется.

VLR обеспечивает также присвоение номера "блуждающей" подвижной станции (MSRN). Когда подвижная станция принимает входящий вызов, VLR выбирает его MSRN и передает его на MSC, который осуществляет маршрутизацию этого вызова к базовым станциям, находящимся рядом с подвижным абонентом.

VLR также распределяет номера передачи управления при передаче соединений от одного MSC к другому. Кроме того, VLR управляет распределением новых временных международных идентификационных номеров (TMSI) и передает их в HLR. Он также управляет процедурами установления подлинности во время обработки вызова. По решению оператора TMSI может периодически изменяться для усложнения процедуры идентификации абонентов. Доступ к базе данных VLR может обеспечиваться через IMSI, TMSI или MSRN. В целом VLR представляет собой локальную базу данных о подвижном абоненте для той зоны, где находится абонент, что позволяет исключить постоянные запросы в HLR и сократить время на обслуживание вызовов.

Для исключения несанкционированного использования ресурсов системы связи вводятся механизмы аутентификации - удостоверения подлинности абонента. Центр аутентификации состоит из нескольких блоков и формирует ключи и алгоритмы аутентификации. С его помощью проверяются полномочия абонента и осуществляется его доступ к сети связи. AUC принимает решения о параметрах процесса аутентификации и определяет ключи шифрования абонентских станций на основе базы данных, сосредоточенной в регистре идентификации оборудования (EIR - Equipment Identification Register).

Каждый подвижный абонент на время пользования системой связи получает стандартный модуль подлинности абонента (SIM), который содержит: международный идентификационный номер (IMSI), свой индивидуальный ключ аутентификации (Ki), алгоритм аутентификации (A3).

С помощью записанной в SIM информации в результате взаимного обмена данными между подвижной станцией и сетью осуществляется полный цикл аутентификации и разрешается доступ абонента к сети.

Процедура проверки сетью подлинности абонента реализуется следующим образом. Сеть передает случайный номер (RAND) на подвижную станцию. На ней с помощью Ki и алгоритма аутентификации A3 определяется значение отклика (SRES), т.е.

$$SRES = Ki * [RAND]$$

Подвижная станция посылает вычисленное значение SRES в сеть, которая сверяет значение принятого SRES со значением SRES, вычисленным сетью. Если оба значения совпадают, подвижная станция приступает к передаче сообщений. В противном случае связь прерывается, и индикатор подвижной станции показывает, что опознавание не состоялось. Для обеспечения секретности вычисление SRES происходит в рамках SIM. Несекретная информация (например, Ki) не подвергается обработке в модуле SIM.

EIR - регистр идентификации оборудования, содержит централизованную базу данных для подтверждения подлинности международного идентификационного номера оборудова-

ния подвижной станции (IMEI). Эта база данных относится исключительно к оборудованию подвижной станции. База данных EIR состоит из списков номеров IMEI, организованных следующим образом:

БЕЛЫЙ СПИСОК - содержит номера IMSI, о которых есть сведения, что они закреплены за санкционированными подвижными станциями.

ЧЕРНЫЙ СПИСОК - содержит номера IMSI подвижных станций, которые украдены или которым отказано в обслуживании по другой причине.

СЕРЫЙ СПИСОК - содержит номера IMSI подвижных станций, у которых существуют проблемы, выявленные по данным программного обеспечения, что не является основанием для внесения в "черный список".

ПБС (BSS) - оборудование базовой станции, состоит из контроллера базовой станции КБС (BSC) и приемо-передающих базовых станций БС (BTS). Контроллер базовой станции может управлять несколькими приемо-передающими блоками. BSS управляет распределением радиоканалов, контролирует соединения, регулирует их очередность, обеспечивает режим работы с прыгающей частотой, модуляцию и демодуляцию сигналов, кодирование и декодирование сообщений, кодирование речи, адаптацию скорости передачи для речи, данных и вызова, определяет очередность передачи сообщений персонального вызова.

К числу операций, производимых КБС, относятся передача обслуживания при переходе МС между сотами, контролируемые одним и тем же КБС, передача вызовов между МС, находящимися в зонах действия разных БС, но одного и того же контроллера (в противном случае, первичное управление возлагается на ЦКМС). Под управлением КБС осуществляются радиоизмерения в каналах связи, регулируется мощность передатчиков мобильных и базовых станций.

Внутренние GSM – интерфейсы

Интерфейс между MSC и HLR совмещен с VLR (В-интерфейс). Когда MSC необходимо определить местоположение подвижной станции, он обращается к VLR. Если подвижная станция инициирует процедуру местоопределения с MSC, он информирует свой VLR, который заносит всю изменяющуюся информацию в свои регистры. Эта процедура происходит всегда, когда MS переходит из одной области местоопределения в другую. В случае, если абонент запрашивает специальные дополнительные услуги или изменяет некоторые свои данные, MSC также информирует VLR, который регистрирует изменения и при необходимости сообщает о них HLR.

Интерфейс между ЦКМС (MSC) и HLR (С-интерфейс) используется для обеспечения взаимодействия между MSC и HLR. MSC может послать указание (сообщение) HLR в конце сеанса связи для того, чтобы абонент мог оплатить разговор. Когда сеть фиксированной телефонной связи не способна исполнить процедуру установления вызова подвижного абонента, MSC может запросить HLR с целью определения местоположения абонента для того, чтобы послать вызов MS.

Интерфейс между HLR и VLR (D-интерфейс) используется для расширения обмена данными о положении подвижной станции, управления процессом связи. Основные услуги, предоставляемые подвижному абоненту, заключаются в возможности передавать или принимать сообщения независимо от местоположения. Для этого HLR должен пополнять свои данные. VLR сообщает HLR о положении MS, управляя ею и переприсваивая ей номера в процессе блуждания, посылает все необходимые данные для обеспечения обслуживания подвижной станции.

Интерфейс между MSC (Е-интерфейс) обеспечивает взаимодействие между разными MSC при осуществлении процедуры HANDOVER - "передачи" абонента из зоны в зону при его движении в процессе сеанса связи без ее перерыва.

4.3.7 Структура TDMA кадров и формирование сигналов в стандарте GSM

В результате анализа различных вариантов построения цифровых сотовых систем подвижной связи (ССПС) в стандарте GSM принят многостанционный доступ с временным разделением каналов (TDMA). Общая структура временных кадров показана на рис. 4.7. Длина периода последовательности в этой структуре, которая называется гиперкадром, равна $T_g = 3 \text{ ч } 28 \text{ мин } 53 \text{ с } 760 \text{ мс}$ ($12533,76 \text{ с}$). Гиперкадр делится на 2048 суперкадров, каждый из которых имеет длительность $T_e = 12533,76/2048 = 6,12 \text{ с}$.

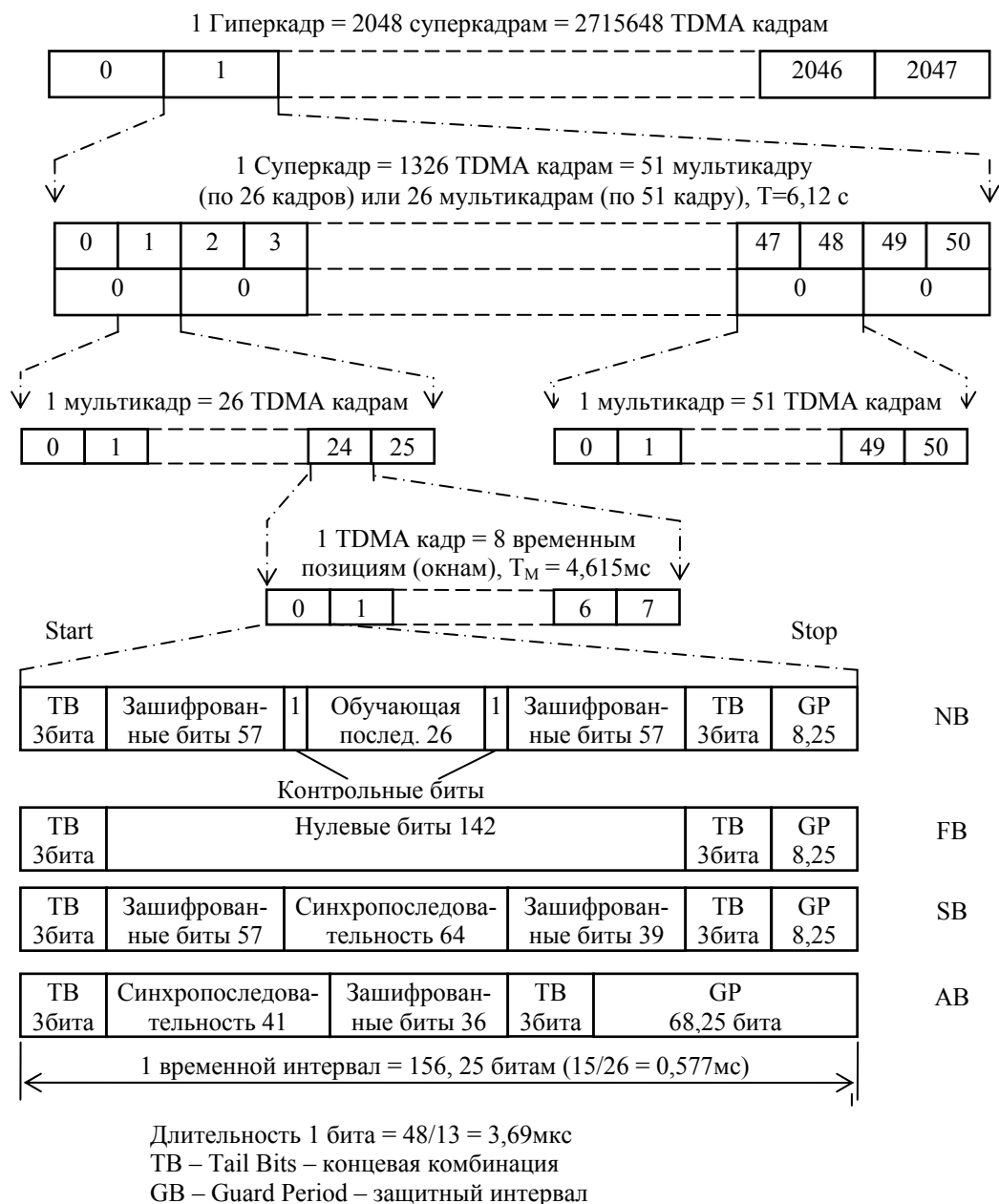


Рис. 4.7. Временная структура кадров стандарта GSM

Таким образом, гиперкадр состоит из 2715647 TDMA кадров. Необходимость такого большого периода гиперкадра объясняется требованиями применяемого процесса криптографической защиты, в котором номер кадра NF используется как входной параметр. TDMA кадр делится на восемь временных позиций с периодом

$$T_o = 60/13:8 = 576,9 \text{ мкс } (15/26 \text{ мс})$$

Каждая временная позиция обозначается TN с номером от 0 до 7. Физический смысл временных позиций, которые иначе называются окнами, - время, в течение которого осуществляется модуляция несущей цифровым информационным потоком, соответствующим речевому сообщению или данным.

Цифровой информационный поток представляет собой последовательность пакетов, размещаемых в этих временных интервалах (окнах). Пакеты формируются немного короче, чем интервалы, их длительность составляет 0,546 мс, что необходимо для приема сообщения при наличии временной дисперсии в канале распространения.

Информационное сообщение передается по радиоканалу со скоростью 270,833 кбит/с.

Это означает, что временной интервал TDMA кадра содержит 156,25 бит.

Длительность одного информационного бита $576,9 \text{ мкс} / 156,25 = 3,69 \text{ мкс}$.

Каждый временной интервал, соответствующий длительности бита, обозначается BN с номером от 0 до 155; последнему интервалу длительностью 1/4 бита присвоен номер 156.

Для передачи информации по каналам связи и управления, подстройки несущих частот, обеспечения временной синхронизации и доступа к каналу связи в структуре TDMA кадра используются пять видов временных интервалов (окон):

NB используется для передачи информации по каналам связи и управления, за исключением канала доступа RACH. Он состоит из 114 бит зашифрованного сообщения и включает защитный интервал (GP) в 8,25 бит длительностью 30,46 мкс. Информационный блок 114 бит разбит на два самостоятельных блока по 57 бит, разделенных между собой обучающей последовательностью в 26 бит, которая используется для установки эквалайзера в приемнике в соответствии с характеристиками канала связи в данный момент времени. Под эквалайзером здесь понимается трансверсальный фильтр, согласованный с каналом (см. раздел 3.2.1.5).

В состав NB включены два контрольных бита (Stealing Flag), которые служат признаком того, содержит ли передаваемая группа речевую информацию или информацию сигнализации. В последнем случае информационный канал (Traffic Channel) "уведен" для обеспечения сигнализации.

Между двумя группами зашифрованных бит в составе NB находится обучающая последовательность из 26 бит, известная в приемнике. С помощью этой последовательности обеспечивается:

- оценка частоты появления ошибок в двоичных разрядах по результатам сравнения принятой и эталонной последовательностей. В процессе сравнения вычисляется параметр RXQUAL, принятый для оценки качества связи. Конечно, речь идет только об оценке связи, а не о точных измерениях, так как проверяется только часть передаваемой информации. Параметр RXQUAL используется при вхождении в связь, при выполнении процедуры "эстафетной передачи" (Handover) и при оценке зоны покрытия радиосвязью;

- оценка импульсной характеристики радиоканала на интервале передачи NB для последующей коррекции тракта приема сигнала за счет использования адаптивного эквалайзера в тракте приема;

- определение задержек распространения сигнала между базовой и подвижной станциями для оценки дальности связи. Эта информация необходима для того, чтобы пакеты данных от разных подвижных станций не накладывались при приеме на базовой станции. Поэтому удаленные на большее расстояние подвижные станции должны передавать свои пакеты раньше станций, находящихся в непосредственной близости от базовой станции. FB предназначен для синхронизации по частоте подвижной станции. Все 142 бита в этом временном интервале - нулевые, что соответствует немодулированной несущей со сдвигом 1625/24 кГц выше номинального значения частоты несущей. Это необходимо для проверки работы своего передатчика и приемника при небольшом частотном разносе каналов (200 кГц), что со-

ставляет около 0,022% от номинального значения диапазона частот 900 МГц. FB содержит защитный интервал 8,25 бит так же, как и нормальный временной интервал. Повторяющиеся временные интервалы подстройки частоты (FB) образуют канал установки частоты (FCSN).

SB используется для синхронизации по времени базовой и подвижной станций. Он состоит из синхропоследовательности длительностью 64 бита, несет информацию о номере TDMA кадра и идентификационный код базовой станции. Этот интервал передается вместе с интервалом установки частоты. Повторяющиеся интервалы синхронизации образуют так называемый канал синхронизации (SCH).

DB обеспечивает установление и тестирование канала связи. По своей структуре DB совпадает с NB (рис. 1.6) и содержит установочную последовательность длиной 26 бит. В DB отсутствуют контрольные биты и не передается никакой информации. DB лишь информирует о том, что передатчик функционирует.

AB обеспечивает разрешение доступа подвижной станции к новой базовой станции. AB передается подвижной станцией при запросе канала сигнализации. Это первый передаваемый подвижной станцией пакет, следовательно, время прохождения сигнала еще не измерено. Поэтому пакет имеет специфическую структуру. Сначала передается концевая комбинация 8 бит, затем - последовательность синхронизации для базовой станции (41 бит), что позволяет базовой станции обеспечить правильный прием последующих 36 зашифрованных бит. Интервал содержит большой защитный интервал (68,25 бит, длительностью 252 мкс), что обеспечивает (независимо от времени прохождения сигнала) достаточное временное разнесение от пакетов других подвижных станций,

Этот защитный интервал соответствует двойному значению наибольшей возможной задержки сигнала в рамках одной соты и тем самым устанавливает максимально допустимые размеры соты. Особенность стандарта GSM - возможность обеспечения связью подвижных абонентов в сотах с радиусом около 35 км. Время распространения радиосигнала в прямом и обратном направлениях составляет при этом 233,3 мкс.

Одна из особенностей формирования сигналов в стандарте GSM - использование медленных скачков по частоте в процессе сеанса связи. Главное назначение таких скачков (SFH - Slow Frequency Hopping) - обеспечение частотного разнесения в радиоканалах, функционирующих в условиях многолучевого распространения радиоволн. SFH используется во всех подвижных сетях, что повышает эффективность кодирования и перемежения при медленном движении абонентских станций. Принцип формирования медленных скачков по частоте состоит в том, что сообщение, передаваемое в выделенном абоненту временном интервале TDMA кадра (577 мкс), в каждом последующем кадре передается (принимается) на новой фиксированной частоте. В соответствии со структурой кадров время для перестройки частоты составляет около 1 мс.

В процессе скачков по частоте постоянно сохраняется дуплексный разнос 45 МГц между каналами приема и передачи. Всем активным абонентам, находящимся в одной соте, ставятся в соответствие ортогональные формирующие последовательности, что исключает взаимные помехи при приеме сообщений абонентами в соте. Параметры последовательности переключения частот (частотно-временная матрица и начальная частота) назначаются каждой подвижной станции в процессе установления канала. Ортогональность последовательностей переключения частот в соте обеспечивается начальным частотным сдвигом одной и той же (по алгоритму формирования) последовательности. В смежных сотах используются различные формирующие последовательности.

Принятая структура TDMA кадров и принципы формирования сигналов в стандарте GSM в совокупности с методами канального кодирования позволили снизить требуемое для

приема отношение сигнал/помеха до 9 дБ, тогда как в стандартах аналоговых сотовых сетей связи оно составляет 17-18 дБ.

4.4 Критерии выбора модуляционных форматов при цифровой передаче данных

Как и вообще в инженерном проектировании, при предпочтении того или иного способа модуляции приходится искать компромисс между конфликтующими требованиями. В первую очередь речь идет о разумном использовании радиочастотного ресурса: с одной стороны, избранный вид модуляции должен обеспечивать необходимую достоверность передачи, с другой – обладать достаточной спектральной эффективностью, т. е. распоряжаться выделенным участком диапазона с должной экономностью. Противоречивость этих требований очевидна: как следует из фундаментальных положений теории связи, при прочих равных условиях качество передачи возрастает с расширением спектра сигнала, тогда как стремление улучшить показатели абонентской емкости (см. § 5.4), электромагнитной совместимости и удельной (приходящейся на единицу полосы) скорости передачи, наоборот, подталкивает к бережливости в расходовании предоставленного частотного ресурса.

Обсудим подходы к выбору дискретных форматов модуляции (манипуляции), характерные для цифровых стандартов мобильной связи второго и третьего поколений. В качестве общей модели сигнала примем последовательность одинаковых по форме импульсов (посылок), повторяющихся с постоянным интервалом T :

$$s(t) = \sum_{i=-\infty}^{\infty} S_0(t - iT) \cos(2\pi f_0 t + \varphi_i), \quad (4.1)$$

где $S_0(t)$ – огибающая посылки; φ_i – начальная фаза i -й посылки.

В простейшем случае *бинарной фазовой манипуляции* – БФМ, которая реализует наиболее помехоустойчивый способ двоичной телеграфии, в отношении расходования частотного ресурса она, однако, оказывается чрезвычайно неэффективной. Дело в том, что спектр мощности сигнала (4.1) совпадает по форме с энергетическим спектром посылки $S_0(t)$, и когда последняя прямоугольна, убывает с частотой f весьма медленно – пропорционально $1/f^2$. Если, как это часто делается в технике связи при регламентации использования радиоспектра, измерять полосу сигнала шириной окна Δf_{99} , в котором удерживается не менее 99% излучаемой мощности, для БФМ получится цифра, многократно превосходящая традиционный ориентир $1/T$: $\Delta f_{99} \approx 18.5/T$. По этой причине БФМ с прямоугольными посылками не применяется в цифровой мобильной телефонии.

Для повышения спектральной эффективности можно наметить несколько путей. Простейший из них состоит в увеличении длительности прямоугольной посылки T с сохранением прежней скорости передачи R_t в числе бит на единицу времени. При БФМ один бит передается за время $T_b = T$, так что $R_t = 1/T_b$. Для удержания этой скорости при "удлинении" посылки следует увеличить число возможных значений фазы φ_i . Так, при удвоении длительности посылки ($T = 2T_b$) в течение временного отрезка T придется передавать 2 бита информации, т. е. 4 различных сообщения, что может быть достигнуто за счет использования четырех разрешенных значений фазы вместо двух, например $0, \pi, \pi/2, -\pi/2$. Такой способ манипуляции называют *квадратурной ФМ* – КФМ (QPSK – *quadrature phase shift keying*). Можно видеть, что КФМ в 2 раза экономнее БФМ в отношении использования частотного ресурса, поскольку имеет спектр той же формы, но суженный вдвое за счет двукратного растяжения посылки. Особо подчеркнем, что указанный выигрыш достигнут без ухудшения помехоустойчивости приема. В самом деле, пусть энергия посылки при БФМ равна E_b . Тогда евклидово расстояние между противоположными посылками (геометрически представляющими собой

противоположные векторы длины $\sqrt{E_b}$), определяющее вероятность их перепутывания, составит $2\sqrt{E_b}$ (рис. 4.8, а).

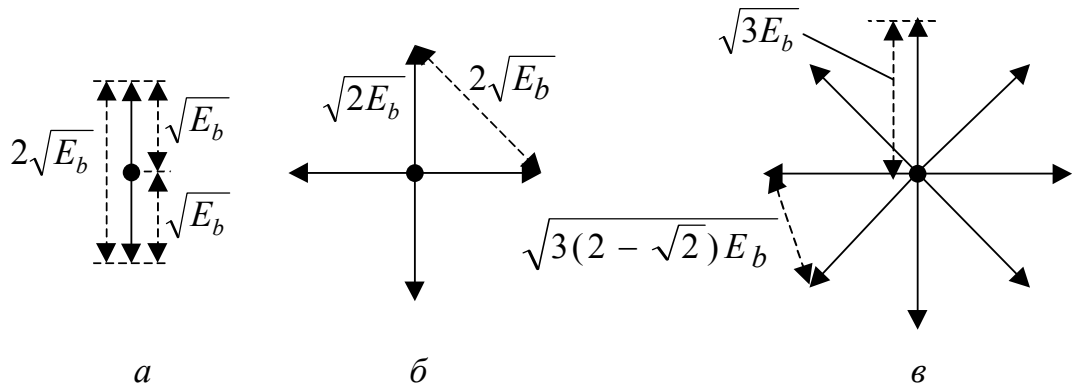


Рис. 4.8. Геометрическая интерпретация фазовой манипуляции

При КФМ четырем сообщениям отвечает четверка биортогональных векторов длины $\sqrt{2E_b}$ (рис. 4.8, б), так как при неизменной мощности энергии посылки E_q удвоится по сравнению со случаем БФМ за счет удвоения длительности: $E_q = 2E_b$. При этом расстояние между соседними векторами, определяющее наибольшую из вероятностей перепутывания сообщений, останется прежним ($\sqrt{2E_q} = 2\sqrt{E_b}$), что и означает отсутствие сколько-нибудь заметного ухудшения помехоустойчивости приема при переходе от БФМ к КФМ.

Не составляет труда убедиться, что при дальнейшем увеличении длительности посылки требование поддержания постоянства скорости приведет к сближению соседних векторов. Так, утроение длительности без снижения скорости означает передачу одной посылкой восьми сообщений, так что трехкратный рост энергии посылки по сравнению с БФМ будет нивелироваться уменьшением угла между соседними сигналами до 45° (рис. 4.8, в), т.е. уменьшением минимального евклидова расстояния до $\sqrt{3E_b(2-\sqrt{2})}$. Таким образом, в данном случае трехкратный выигрыш в полосе приобретается ценой энергетических потерь порядка 3,5 дБ (именно таким должно быть увеличение энергии, компенсирующее сближение векторов и снижающее вероятность ошибки до прежнего уровня). Дальнейшее повышение спектральной эффективности подобным способом окажется тем более невыгодным в плане энергозатрат: M -кратный выигрыш в полосе при 2^M -ичной ФМ будет сопровождаться энергетическим проигрышем в

$$\gamma = \frac{2}{M(1 - \cos \frac{\pi}{2^{M-1}})}$$

раз.

В принципе существует возможность определенного снижения упомянутых энергетических потерь за счет оптимизации созвездия сигнальных векторов на плоскости, максимизирующей минимальное расстояние между сигнальными векторами. При этом последние имеют неодинаковую длину, т.е. фазовая манипуляция дополняется параллельной амплитудной. Подобные форматы, известные под названиями амплитудно-фазовая и квадратурная амплитудная манипуляции (АФМ и КАМ), широко распространены во многих телекоммуникационных сетях (кабельная, радиорелейная связь и т. д.). Однако, специфика систем беспроводной мобильной телефонии состоит в исключительной важности эффективного энергосбережения, продлевающего срок автономной (без подзарядки или смены батарей) работы портативного терминала и способствующего коммерческой привлекательности его массогабаритных характеристик. По этим причинам многократная (с числом фаз 16 и более) ФМ,

наряду с АФМ и КАМ, не рассматриваются как приемлемые альтернативы для организации радиоинтерфейсов сетей мобильной связи, хотя восьмиуровневая ФМ (8-PSK) избрана как инструмент увеличения скорости передачи в системах второго поколения в рамках спецификации EDGE (см. раздел 4.7).

4.5 Модуляционные форматы цифровых стандартов сотовой связи второго и третьего поколений

Различные модификации КФМ находят самое широкое применение в сетях цифровой мобильной связи как второго, так и третьего поколений. Мотивами усовершенствований по сравнению с базовым вариантом КФМ служат, с одной стороны, все та же заинтересованность в компактном спектре, а с другой – стремление к оптимизации энергетического режима усилителя мощности передатчика подвижного терминала. Поясним подробнее смысл последнего фактора.

Энергопотребление передатчика в значительной мере зависит от режима окончного усилителя мощности или, иначе говоря, от пик-фактора сигнала, равного отношению его пиковой и средней мощностей. Наиболее благоприятным считается режим класса С, когда активный элемент усилителя постоянно находится вблизи точки насыщения, выполняя, по существу, функции ключа, и требования к его линейному динамическому диапазону минимальны. Для приближения к подобному режиму передаваемый сигнал должен быть свободен от глубокой амплитудной модуляции, т.е. иметь пик-фактор, близкий к единице. Данное условие, в свою очередь, означает, что скачки мгновенной фазы, сопровождающие переход от данной посылки к последующей, должны быть по возможности минимизированы.

Обратимся за подтверждением к рис. 4.9. Когда значения передаваемых символов 0 и 1 при БФМ (рис. 4.9, а) равновероятны, в среднем в половине переходов между последовательными посылками наблюдается скачок фазы на 180° , т.е. смена полярности на противоположную. Изменение мгновенной комплексной огибающей от A до $-A$ или наоборот должно быть без искажений передано усилителем мощности, которому тем самым навязывается чрезмерное требование в части линейного динамического диапазона ($2A$). Поскольку при линейной передаче перепадов большая часть мощности рассеивается на активном элементе усилителя, становится понятным желание снизить частоту и размах скачков фазы в манипулированном сигнале. Переход от БФМ к стандартной КФМ (рис 4.9, б) с сохранением прежней мощности не ослабляет требований к динамическому диапазону усилителя, поскольку максимальные скачки фазы в 180° сохраняются, хотя и случаются значительно (примерно вдвое) реже.

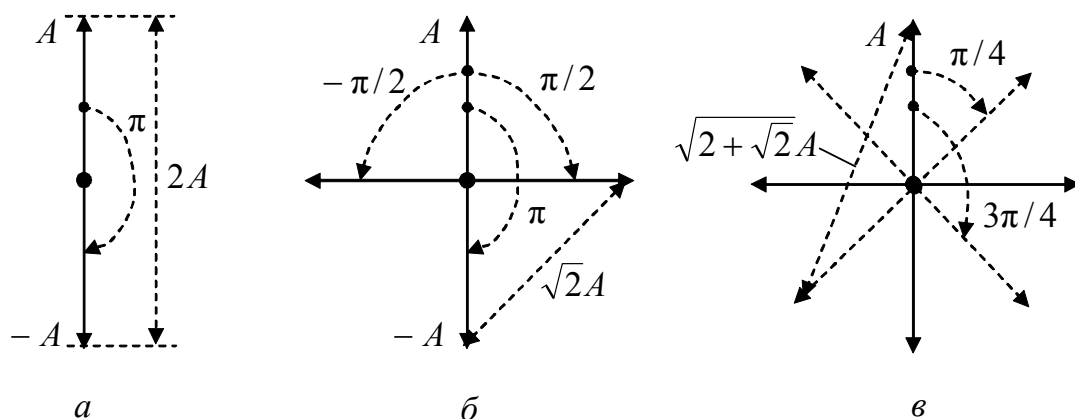


Рис. 4.9. Перепады мгновенной амплитуды при различных видах ФМ

Смягчение требований к динамическому диапазону передатчика достигается в формате КФМ со сдвигом (OQPSK – *offset* QPSK), для понимания смысла которого разумно сначала ввести в (4.1) начальную фазу в $\pi/4$ (разумеется, это никак не повлияет на свойства КФМ и лишь повернет целиком диаграмму рис. 4.9, б на 45°), а затем переписать результат в форме:

$$s(t) = \sum_{i=-\infty}^{\infty} S_0(t-iT) \cos(2\pi f_0 t + \varphi_i + \frac{\pi}{4}) =$$

$$= \frac{1}{\sqrt{2}} \left[\sum_{i=-\infty}^{\infty} a_i S_0(t-iT) \cos(2\pi f_0 t) + \sum_{i=-\infty}^{\infty} b_i S_0(t-iT) \sin(2\pi f_0 t) \right],$$

где $a_i = \cos \varphi_i - \sin \varphi_i$ и $b_i = -\cos \varphi_i - \sin \varphi_i$ – бинарные символы, принимающие значения +1 и –1. Как видно, КФМ сигнал можно интерпретировать как сумму двух квадратурных БФМ сигналов, что дает возможность демодулировать каждую из последовательностей символов $\{a_i\}$, $\{b_i\}$ отдельно от другой, поскольку синусная и косинусная квадратурные компоненты ортогональны и не создают друг другу перекрестных помех при когерентном приеме. Сдвинем теперь синусную квадратурную составляющую во времени на половину длительности посылки, придя к результату

$$s(t) = \frac{1}{\sqrt{2}} \left[\sum_{i=-\infty}^{\infty} a_i S_0(t-iT) \cos(2\pi f_0 t) + \sum_{i=-\infty}^{\infty} b_i S_0(t-iT - \frac{T}{2}) \sin(2\pi f_0 t) \right], \quad (4.2)$$

сохраняющему ортогональность квадратурных компонент и, следовательно, все характеристики обычной КФМ в части качества передачи данных. Теперь, однако, изменения бинарных символов, манипулирующих квадратурные несущие, происходят неодновременно: в момент смены символа a_i символ другой квадратуры b_i остается неизменным и наоборот. Благодаря этому при каждой смене символа a_i или b_i сигнальный вектор на рис. 4.9,б может перейти только в соседний, но никак не в противоположный. Тем самым требуемый линейный динамический диапазон $\sqrt{2}A$ оказывается в $\sqrt{2}$ раз меньшим по сравнению со случаем БФМ или стандартной КФМ. Именно по этой причине КФМС была избрана для построения обратного (от МС к БС) канала в CDMA стандарте второго поколения IS-95.

Другой вариант достижения той же цели реализован в цифровых стандартах D-AMPS (США, известен также под названиями IS-136 и US-TDMA) и PDC (Япония, прежнее наименование JDC) [13], где вместо сдвига посылок введен поворот на угол $\pi/4$ алфавита значений фаз при переходе от четных посылок к нечетным. Благодаря такому смещению при $i = 2k$, $k = \dots, -1, 0, 1, \dots$ φ_i в (4.1) принимает значения из множества $0, \pi, \pm\pi/2$, а при $i = 2k + 1$ – из множества $\pm\pi/4, \pm3\pi/4$ (пунктир на рис. 4.9, в). Разумеется, на приемной стороне этот сдвиг легко учитывается, так что демодуляция выполняется практически так же, как при обычной КФМ. Подобный вид КФМ получил название $\pi/4$ -КФМ ($\pi/4$ -QPSK). Его преимущество в сравнении с КФМС состоит именно в отсутствии усложнений демодулятора, свойственных КФМС, хотя в отношении смягчения требований к линейному динамическому диапазону $\pi/4$ -КФМ не столь эффективна: переходам на угол $\pm3\pi/4$ соответствует перепад комплексной огибающей, равный $\sqrt{2} + \sqrt{2}A$ и примерно в 1,31 раза превышающий аналогичный показатель КФМС.

Рассмотренные трансформации КФМ никоим образом не влияют на спектральные характеристики модулированного сигнала, которые в рамках случайной модели передаваемого потока данных определяются исключительно формой посылки. Поскольку причиной медленного спада спектра мощности сигналов (4.1) и (4.2) с частотой является разрывность прямоугольной посылки, радикального сужения спектра можно добиться только за счет сглаживания ее формы. Если, в частности, принять в (4.2) за посылку импульс с огибающей в виде положительной полуволны косинуса (см. рис 4.3) с амплитудой $\sqrt{2}A$

$$S_0(t) = \begin{cases} \sqrt{2}A \cos(\frac{\pi t}{T}), & |t| \leq \frac{T}{2}, \\ 0, & |t| > \frac{T}{2}, \end{cases} \quad (4.3)$$

КФМ преобразуется в свой подвид, известный как *минимальная частотная манипуляция* – МЧМ (MSK – *minimal shift keying*).

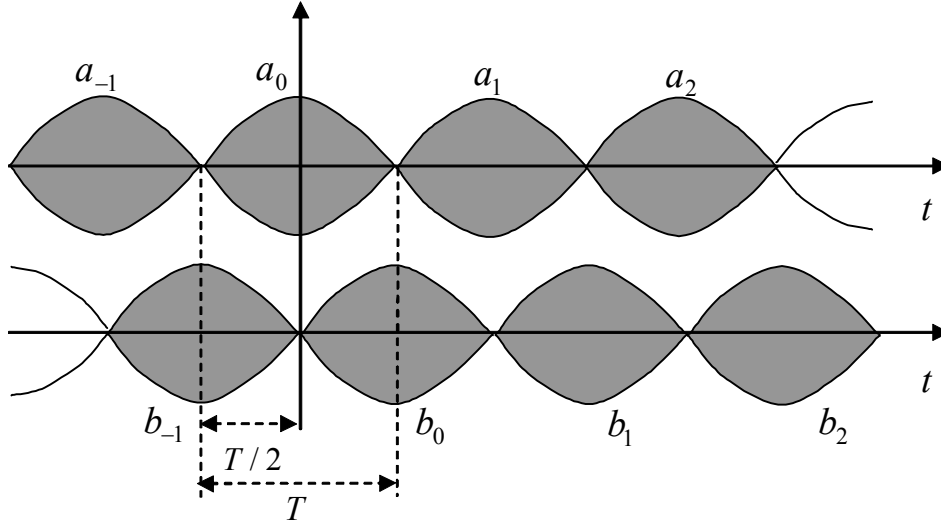


Рис 4.10. Квадратурные составляющие МЧМ сигнала

Для того, чтобы убедиться в преимуществах МЧМ и попутно выяснить корни ее наименования, найдем комплексную огибающую $\dot{S}(t)$ сигнала (4.2) с формой посылки, задаваемой (4.3). При этом в силу полной однородности поведения сигнала во времени достаточно рассмотреть лишь два смежных отрезка $[-T/2, 0]$ и $[0, T/2]$. Используя элементарную тригонометрию, формулу Эйлера и бинарность символов a_i, b_i , в силу которой $a_i/b_i = a_i b_i = \pm 1$, получаем

$$\dot{S}(t) = \begin{cases} a_0 A \cos(\frac{\pi t}{T}) + j b_{-1} A \cos[\frac{\pi}{T}(t + T - \frac{T}{2})] = a_0 A \exp(-j \frac{\pi a_0 b_{-1} t}{T}), & -\frac{T}{2} \leq t < 0, \\ a_0 A \cos(\frac{\pi t}{T}) + j b_0 A \cos[\frac{\pi}{T}(t - \frac{T}{2})] = a_0 A \exp(j \frac{\pi a_0 b_0 t}{T}), & 0 \leq t < \frac{T}{2}, \end{cases} \quad (4.4)$$

Как показывает (4.4), действительная огибающая $S(t) = |\dot{S}(t)| = A$ постоянна и, следовательно, МЧМ не сопровождается амплитудной модуляцией, обеспечивая равенство пик-фактора сигнала единице и, как результат, оптимальность режима усилителя мощности передатчика.

Другой вывод из (4.4) состоит в том, что рассматриваемый вид модуляции сводится, по существу, к бинарной частотной манипуляции прямоугольных посылок длительности $T/2$, поскольку линейное изменение фазы с угловым коэффициентом π/T означает сдвиг частоты на $\pm 1/2T$. Принципиальным является то, что переключение частоты между посылками каждые $T/2$ с происходит без скачков фазы, т.е. МЧМ является модуляцией с непрерывной фазой. В этом легко убедиться, подставив $t = 0$ в обе строки (4.4) и придя к одному и тому же результату:

$$\dot{S}(t = 0^-) = a_0 = \dot{S}(t = 0^+),$$

где верхние индексы "–" и "+" символизируют приближение к точке $t = 0$ соответственно слева и справа.

Таким образом, при любом текущем передаваемом символе очередная посылка начинается с той фазы, которая "набежала" в течение предыдущей. Сказанное иллюстрируется "деревом" траекторий фазы на рис 4.11. В течение каждого отрезка $[kT/2, (k+1)T/2]$ фаза линейно растет или убывает в соответствии с текущим приращением частоты $\pm 1/2T$ (определяемым, в свою очередь, комбинацией символов a_i, b_i сигнала (4.2) на данном отрезке). Значения фазового угла в момент $(k+1)T/2$, отвечающие двум возможным частотам, отличаются ровно на π . При этом любая из возможных траекторий фазы оказывается непрерывной функцией.

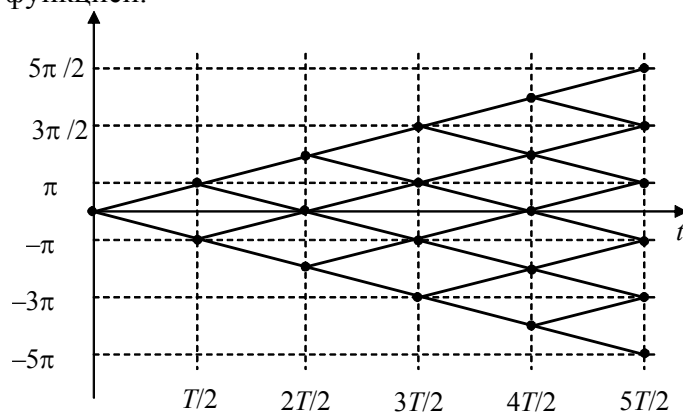


Рис. 4.11. Траектории мгновенной фазы МЧМ сигнала

Отсутствие разрывов фазы обеспечивает спектру МЧМ значительно большую степень компактности по сравнению со стандартной КФМС: спектр мощности МЧМ сигнала убывает примерно пропорционально $1/f^4$, так что занимаемая им полоса сокращается более чем на порядок:

$$\Delta f_{99} \approx \frac{1.2}{T_b}.$$

Резервы дальнейшего сжатия спектра модулированного сигнала кроются в устранении разрывов не только самой фазы, но и ее производных (частоты, скорости изменения частоты и т.д.). Иными словами, линейно-ломаные траектории на рис. 4.11 могут быть заменены более гладкими. В стандарте GSM применен именно такой вариант модуляции – *гауссовская* МЧМ (GMSK – *Gaussian* MSK), при которой закон изменения фазы в течение посылки повторяет ход гауссовской интегральной функции распределения, чем обеспечивается плавность изменения фазы и частоты, а значит, высокая степень компактности спектра. Технически гауссовская (как и обычная) МСК может быть реализована разными средствами, однако для объяснения смысла ее параметров согласно спецификации GSM разумно считать, что исходный поток битовых прямоугольных посылок длительности T_b пропускается через низкочастотный фильтр с гауссовской амплитудно-частотной характеристикой и полосой B (на уровне -3 дБ), после чего сглаженный сигнал модулирует частоту задающего генератора. В стандарте жестко зафиксировано значение $BT_b = 0.3$, отвечающее полосе

$$\Delta f_{99} \approx \frac{0.92}{T_b}.$$

С упрощениями, игнорирующими второстепенные детали, дерево траекторий фазы гауссовской МЧМ показано на рис. 4.12.

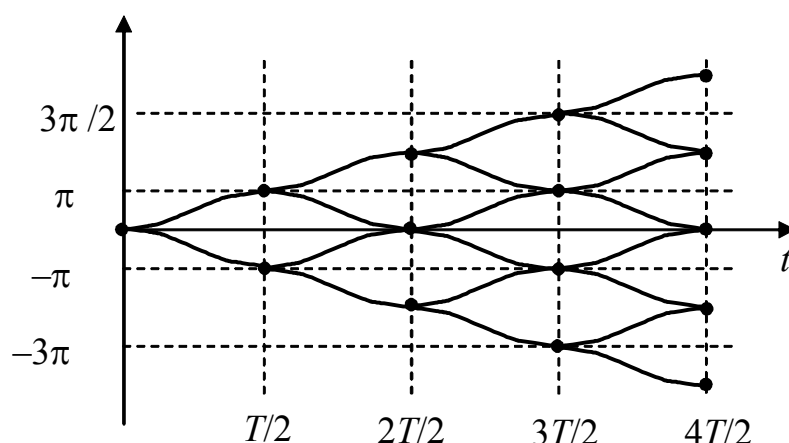


Рис. 4.12. Траектории фазы при гауссовской МЧМ

Отметим, что в эволюции видов модуляции от БФМ до МЧМ обмен между спектральной эффективностью и достоверностью передачи фактически отсутствовал: повышение спектральной эффективности не сопровождалось снижением минимального евклидова расстояния в сигнальном созвездии, что было продемонстрировано на примере сравнения КФМ с БФМ. Случай гауссовской МЧМ несколько выпадает из этого ряда в связи с тем, что неограниченность отклика гауссовского фильтра во времени не позволяет, строго говоря, трактовать данный вид модуляции как сглаженную версию КФМС. В литературе можно встретить упоминание об энергетическом проигрыше гауссовской МЧМ относительно БФМ в 0,46 дБ, однако не исключено, что ее теоретический потенциал выше с учетом возможности демодуляции сообщения "в целом", т.е. на интервале, охватывающем несколько посылок (например, с помощью алгоритма Витерби).

Приведем итоговую таблицу, связывающую ширину полосы Δf_{99} модулированного сигнала со скоростью передачи данных R_t (бит/с) для рассмотренных форматов.

Таблица 4.1. Соотношения между полосой и скоростью передачи для различных видов модуляции

Вид модуляции	Б ФМ	КФМ, КФМС, $\pi/4$ -КФ М,	М ЧМ	гауссов- ская МЧМ
$\Delta f_{99} / R_t$	1 8,5	9,2	1, 2	0,92

Как показывает приведенный обзор, даже в относительно узких рамках цифровых стандартов сотового радиотелефона применяемые методы модуляции отличаются заметным разнообразием. Их перечень станет еще более обширным, если обратиться и к другим телекоммуникационным системам (модемной, радиорелейной, спутниковой связи, персонального вызова, бесшнурового телефона и т. п.). Все это лишний раз свидетельствует о неоднозначности и многовариантности задач идеологического проектирования и существовании параллельных путей достижения оптимальных системных показателей.

4.6 Радиоинтерфейс системы cdmaOne (IS-95)

4.6.1 Общая характеристика системы CDMAOne

С опорой на преимущества, предоставляемые технологией CDMA, компанией Qualcomm (США) была разработана ССМС общего пользования с кодовым разделением каналов. При этом основными стимулами явились увеличение абонентской емкости по сравнению с действующими ССМС, основанными на традиционном частотно/временном доступе, улучшение качества обслуживания и повышение информационной безопасности. Технические требования к системе сформулированы в ряде стандартов Ассоциации Промышленности связи (*Telecommunications Industry Association – TIA*), определяющих характер и порядок взаимодействия различных функциональных узлов ССМС и условия совместимости аппаратуры различных фирм–производителей.

Современные сети, базирующиеся на стандарте IS-95, обеспечивают передачу сигнала со скоростью 9,6 кбит/с (с кодированием) и 14,4 кбит/с (без кодирования), тогда как исходные спецификации cdmaOne предполагали скорость передачи 8 кбит/с, 13 кбит/с и 8 кбит/с EVRC (Enhanced Variable Rate Vocoder). В настоящее время широко применяется версия стандарта **IS-95A**.

Версия IS-95B основана на объединении нескольких каналов CDMA, организуемых в прямом направлении (от базовой станции к мобильной). Скорость может увеличиваться до 28,8 кбит/с (при объединении двух каналов по 14,4 кбит/с) или до 115,2 кбит/с (8 каналов по 14,4 кбит/с). Сети на основе IS-95B смогут обеспечивать доступ в Internet до появления систем третьего поколения. Однако для того, чтобы предоставлять услуги пакетной передачи, контроллер БС нужно дооснастить маршрутизатором. В спецификациях стандарта предусмотрено качественное улучшение характеристик обслуживания за счет снижения потерь при переходе абонента от одной БС к другой, а также повышение точности контроля мощности до 0,25 дБ, организация каналов приоритетного доступа и другие усовершенствования.

В версии IS-95C (CDMA 2000) модификации коснулись повышения частотной эффективности и увеличения емкости телефонной сети в два раза. Спецификациями предусматривается дополнительный канал с ортогональным сдвигом несущей, по которому может передаваться полный кодовый ансамбль сигналов (т. е. 64 кода Уолша), такой же, как и по синфазному каналу. Системы на базе **CDMA 2000** обратно совместимы с сетями на основе IS-95A и **IS-95B** и сохраняют прежнюю полосу частот – 1,25 МГц. По сравнению с предыдущими версиями скорость передачи в системе возрастает до 144 кбит/с; при этом сокращается энергопотребление терминала.

К числу основных стандартов рассматриваемой ССМС относятся:

IS-95 – эфирный интерфейс (радиоинтерфейс);

IS-96 – интерфейс речевых служб;

IS-97 – интерфейс мобильной станции;

IS-98 – интерфейс базовой станции;

IS-99 – интерфейс службы передачи данных.

Учитывая особую значимость радиоинтерфейса, рассматриваемую систему называют ССМС стандарта IS-95, либо, принимая во внимание технологию организации множественного доступа, системой cdmaOne.

Система IS-95 рассчитана на работу в диапазоне частот 800 МГц, причем для прямого канала (линия "вниз") выделен участок спектра 869,04...893,97 МГц, а для обратного (линия "вверх") – 824,04...848,96 МГц. Ширина полосы канала связи составляет 1,25 МГц, поэтому при развертывании ССМС IS-95 операторы могут осуществлять частотное планирование исходя из указанных полос. Однако согласно решению Федеральной Комиссии связи США, одному оператору может быть выделен максимальный диапазон частот, равный 12,5 МГц, как в прямом (от БС к МС), так и в обратном (от МС к БС) направлениях, что соответствует 10 физическим частотным радиоканалам с полосой в 1,25 МГц.

Как указывалось в разделе 2.5, технология организации множественного доступа с кодовым разделением каналов основана на применении сложных (*spread spectrum*) сигналов, полоса которых значительно превышает ширину спектра информационного сообщения. В

системе IS-95 реализовано прямое расширение спектра (DSSS) с использованием функций Уолша длины 64 и двух типов псевдослучайных последовательностей (ПСП): короткой и длинной (таблица 4.3). Линия "вниз" организована на основе синхронного варианта CDMA, использующего для разделения физических абонентских каналов ансамбль сигнатур в виде функций Уолша. Асинхронный режим CDMA в обратном канале осуществлен приписыванием различным абонентам одной соты специфических сдвигов длинной ПСП.

Согласованные фильтры приемников БС квазиоптимальны в условиях взаимной интерференции между абонентами одной соты и весьма чувствительны к неодинаковым уровням сигналов от МС, находящимся на различных расстояниях от БС (эффект «далеко-близко»). Для максимизации абонентской емкости системы необходимо, чтобы терминалы всех абонентов излучали сигнал такой мощности, которая обеспечила бы одинаковый уровень принимаемых БС сигналов. Чем точнее управление мощностью, тем больше абонентская емкость системы.

Выравнивание мощностей сигналов МС на входе приемника БС является неперенным условием работоспособности линии "вверх" системы CDMA. В стандарте IS-95 применена быстродействующая петля автоматической регулировки мощности сигналов МС, принцип действия которой описан в разделе 4.6.5. Регулировка осуществляется в динамическом диапазоне 84 дБ с шагом $\pm 0,5$ дБ и с периодичностью от нескольких микросекунд до 1,25 мс. Резерв фактора речевой активности также утилизирован в системе посредством использования речепреобразующего устройства (вокодера) с переменной частотой преобразования аналогового речевого сигнала в цифровой. В зависимости от активности абонента вокодер формирует потоки данных со скоростями 8,6; 4; 2 и 0,8 кбит/с.

Требуемое качество передачи данных в системе достигается с помощью довольно мощного канального кодирования, выполняемого в несколько этапов. На первом, предварительном этапе цифровой речевой сигнал с выхода вокодера, структурированный в кадры длительностью в 20 мс, кодируется блоковым циклическим кодом (CRC) для формирования индикатора качества кадра и дополняется "хвостом", необходимым для последующего сверточного кодирования. Введение подобной избыточности увеличивает фактические скорости данных, заменяя первоначальный их набор на 9,6; 4,8; 2,4 и 1,2 кбит/с. Следующим этапом является сверточное кодирование. В прямом канале используется код с длиной кодового ограничения 9 и скоростью $1/2$, тогда как код в обратном канале, обладающем меньшей помехоустойчивостью (мощность передатчика МС весьма мала), имеет вдвое большую избыточность, т.е. скорость $1/3$ при той же длине кодового ограничения. Наконец, на третьем этапе выполняется перемежение информационного потока в кадре для нейтрализации эффекта пакетирования ошибок.

Помимо перемежения для борьбы с быстрыми замираниями в системе использовано и многолучевое разнесение, т.е. приемники на основе алгоритма RAKE (см. раздел 3.2.1.5). Для этого на БС используется минимум четыре, а на МС – три параллельно работающих коррелятора. Помимо этих корреляторов, настраиваемых на определенную задержку, в каждом приемнике имеется еще и сканирующий по задержке канал, позволяющий осуществлять настройку RAKE-каналов на сигналы с наибольшей интенсивностью.

Наличие нескольких параллельных каналов корреляционной обработки позволяет осуществить *мягкую эстафетную передачу*, (*soft handoff* или *soft handover*) при переходе МС из одной соты в другую. В процессе мягкой эстафетной передачи МС может поддерживать соединение с двумя и более БС, выбирая сигнал с большей интенсивностью. Указанная процедура позволяет поддерживать высокое качество связи при переключении МС с одной БС на другую и делает эстафетную передачу практически незаметной для пользователя.

Стандарт IS-95 обеспечивает высокую степень безопасности передаваемых данных за счет их скремблирования выборками из вышеупомянутой длинной псевдослучайной последовательности. Ключ (маска) скремблирования индивидуален для каждой МС и формируется по секретному правилу на базе ее идентификационного номера.

Основные характеристики стандарта IS-95 и технические параметры приемо-передающего оборудования представлены в табл. 4.2.

Для систем с CDMA, и, в частности для CCMC IS-95, характерна определенная асимметрия линий "вниз" и "вверх": в первой из них (от БС) все абонентские сигналы передаются из одной пространственной точки и потому могут быть легко синхронизированы (синхронный вариант CDMA). В направлении же от БС к МС приходится применять асинхронное кодовое разделение (см. § 5.3.1). Отметим, что временные шкалы отдельных БС сети IS-95 также синхронизированы. Для этого каждая БС оборудована приемником радионавигационной системы GPS NAVSTAR. Благодаря единому системному времени, кодовое разделение сигналов различных БС, занимающих один и тот же частотный канал, осуществлено за счет применения различных сдвигов одной и той же короткой ПСП.

Таблица 4.2. Основные характеристики стандарта IS-95

Параметр стандарта	Значение
Диапазон частот передачи, МГц:	
МС	824,040...848,860
БС	869,040...893,970
Относит. нестабильность несущей частоты:	
МС	$\pm 2,5 \cdot 10^{-6}$
БС	$\pm 5 \cdot 10^{-8}$
Вид модуляции несущей частоты:	
МС	OQPSK
БС	QPSK
Ширина спектра излучаемого сигнала, МГц	
по уровню -3 дБ	1,23
по уровню -40 дБ	1,5
Тактовая частота ПСП, МГц	1,2288
Количество каналов на одной несущей:	
МС: канал доступа	1
канал обратного трафика	1
БС: пилотный канал	1
канал синхронизации	1
канал персонального вызова	1...7
канал прямого трафика	61...55
Скорость передачи данных в канале, бит/с:	
канал синхронизации	1200
канал персонального вызова и доступа	4800,9600
канал трафика	1200, 2400, 4800, 9600
Кодирование информации в каналах:	
МС: канал доступа и обратного трафика	Сверточ. код с длиной кодового ограничения 9 и скоростью 1/3.
БС: канал синхронизации, персонального вызова и прямого трафика.	Сверт. код с длиной кодового ограничения 9 и скоростью 1/2.
Требуемое отношение сигнал/шум, дБ	6...7
Макс. эффектив. излучаемая мощность, Вт:	
МС: (3-го, 2-го и 1-го класса)	1,0; 2,5; 6,3
БС	до 50
Точность управления мощностью передатчика, дБ	$\pm 0,5$
Чувствительность приемника, дБм:	
МС	-105
БС	-117

Для модуляции сигнала используется три вида функций: «короткая» и «длинная» ПСП и функции Уолша порядков от 0 до 63. Все они являются общими для базовых и мобильных станций, однако реализуют разные функции (табл. 4.3).

Таблица 4.3. Параметры кодовых последовательностей в стандарте IS-95

Тип сигнала	Длина кода	Выполняемые функции	
		Базовая станция	Мобильная станция
Код Уолша	64	Кодовое уплотнение или разделение 64 каналов CDMA	Помехоустойчивое кодирование
Короткий код	32768	Разделение сигналов базовых станций по величине циклического сдвига	Код с одинаковым фиксированным циклическим сдвигом - как опорный сигнал скремблера
Длинный код	$2^{42}-1$ (4.4×10^{12})	Прореженный длинный код - как опорная последовательность скремблера	Длинный код с разными циклическими сдвигами - как адресная последовательность

4.6.2 Архитектура линии "вниз"

Логические каналы линии "вниз" включают:

- пилотный канал (*pilot channel* – PI);
- канал синхронизации (*synchronization channel* – SYNC);
- канал персонального вызова (*paging channel* – PCH);
- канал прямого трафика (*forward traffic channel* – TCH).

Отображение логических каналов на физические в прямом направлении осуществляется с помощью системы ортогональных функций Уолша длины 64: w_i , $i = 0, 1 \dots 63$, где i – номер функции Уолша. Стандартом IS-95 предусматривается организация одного пилотного канала, одного канала синхронизации, от одного до семи каналов вызова (в зависимости от абонентской нагрузки на БС) и от 55 до 62 каналов прямого трафика, поскольку часть каналов вызова допускается использовать в качестве каналов трафика. Соответствие между логическими и физическими каналами показано на рис. 4.13.

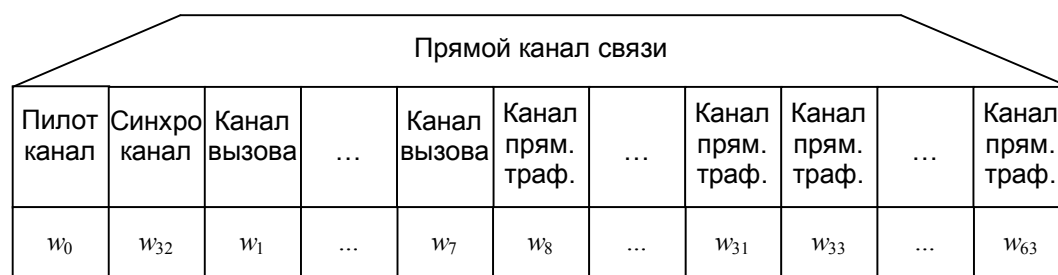


Рис. 4.13. Структура прямого канала CCMC стандарта IS-95

Укрупненная схема преобразований сигнала в передающем тракте базовой станции показана на рис. 4.14:

1. Для различения разных физических каналов данной БС в прямом канале от БС к подвижной – кодирование сигнала функциями Уолша и бинарная фазовая манипуляция.
2. Для шифрования сообщений используется кодирование длинной ПСП и бинарная фазовая манипуляция.
3. Для расширения полосы и различения сигналов разных БС используется кодирование короткой ПСП и квадратурная фазовая манипуляция двумя ПСП одинакового периода.

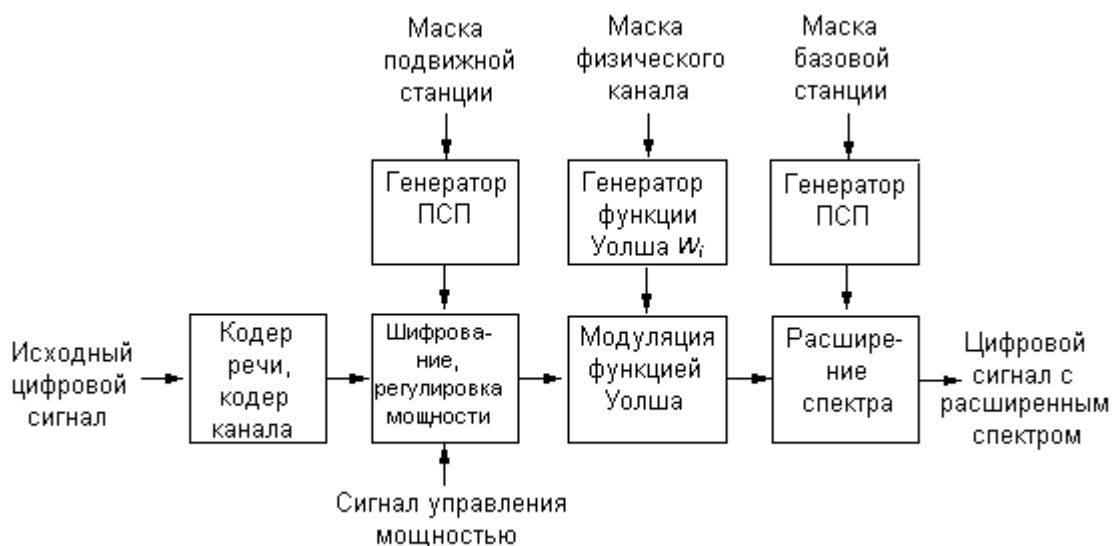


Рис. 4.14. Обработка сигналов в передающем тракте базовой станции

4. Различение сигналов разных станций обеспечивается тем, что все БС используют одну и ту же пару коротких ПСП, но со сдвигом на 64 чипа (дискрета ПСП частотой 1,2288 МГц) между разными станциями. Т.е. всего в сети имеется 511 кодов; при этом все физические каналы одной БС имеют одну и ту же фазу последовательности.

4.6.3 Пилотный канал

Пилотный канал предназначается для начальной синхронизации МС с сетью и оценки параметров прямого канала приемником МС. Мощность, отведенная пилотному каналу, обычно на 4...6 дБ превышает мощность в канале трафика. Тем самым обеспечиваются, с одной стороны, условия для надежного ввода и последующей устойчивой работы петель автоподстройки фазы и задержки когерентного приемника МС, а с другой – точность измерений параметров сигнала БС, достаточная для процедур RAKE, эстафетной передачи и управления мощностью БС.

Согласно рис. 4.13, пилотному каналу присвоена нулевая функция Уолша w_0 , т.е. последовательность из одних нулей. Информационные данные по этому каналу не передаются, т.е. могут также трактоваться как последовательность из одних нулей. После сложения по модулю два этих двух последовательностей результирующий поток (также состоящий из одних нулей)² поступает на преобразователь, превращающий булевы значения символов (0,1) в двуполярные отсчеты (+1,-1) соответственно. После объединения с сигналами других физических каналов (см. ниже) суммарный сигнал поступает на КФМ модулятор, в котором подвергается прямому расширению спектра, перемножаясь с двумя псевдослучайными скремблирующими последовательностями ПСП-I и ПСП-Q (символы I и Q отвечают синфазной и квадратурной составляющим). Период каждой из названных ПСП содержит 2^{15} чипов, частота следования которых согласно стандарту составляет 1,2288 Мчип/с. Прямой подсчет показывает, что на одном двухсекундном отрезке уместается в точности 75 периодов коротких ПСП. Структурно короткие ПСП представляют собой M -последовательности длины $N = 2^{15} - 1$ с характеристическими полиномами

$$f_I = x^{15} + x^{13} + x^9 + x^8 + x^7 + x^5 + 1$$

и

² Тривиальные нулевые последовательности данных и w_0 явно показаны на рис. 4.15, чтобы сохранить единообразие с последующими схемами.

$$f_Q = x^{15} + x^{12} + x^{11} + x^{10} + x^6 + x^5 + x^4 + x^3 + 1,$$

расширенные добавлением нулевого символа к цепочке из 14 последовательных нулей на каждом периоде

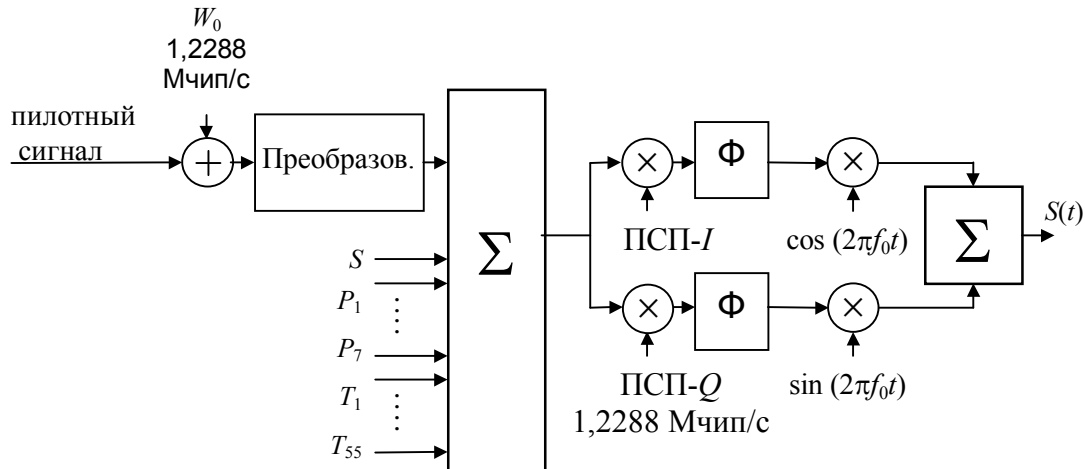


Рис. 4.15. Структурная схема пилотного канала

Как можно видеть, фактически в пилотном канале передается только пара ПСП-I и ПСП-Q или, что эквивалентно комплексная ПСП. Данная комплексная короткая ПСП одинакова для всех 64 CDMA-каналов и используется всеми БС системы, но с разными циклическими сдвигами. Разница в циклических сдвигах позволяет МС разделять сигналы, излучаемые БС разных сот или секторов, т.е. позволяет идентифицировать номер БС либо сектора. Для различных БС сдвиг изменяется с постоянным шагом, равным $64 \text{ чип} \times \text{PILOT_INC}$, где системный параметр PILOT_INC принимает значения от 1 до 4. Таким образом, при минимальном шаге доступны $2^{15} / 2^6 = 512$ сдвигов коротких ПСП, т.е. возможно бесконфликтное существование сети, состоящей из 512 БС. Если же необходимо, чтобы сеть состояла из большего числа БС, то при территориальном планировании сети легко можно добиться, чтобы БС с одинаковыми циклическими сдвигами коротких ПСП не могли одновременно находиться в зоне радиовидимости МС.

С другой стороны, шаг сдвига ПСП однозначно определяет размер соты (или сектора), при котором МС с гарантией различает ПСП, имеющие минимальный временной сдвиг. Нетрудно убедиться, что при минимальном сдвиге в 64 чипа радиус соты составит порядка 15,5 км.

После перемножения с квадратурными ПСП сигнал в каждом из квадратурных плеч КФМ-модулятора фильтруется (фильтры Φ на рис. 4.15) для формирования приемлемого спектра и перемножается со сдвинутыми на 90° гармоническими колебаниями центральной частоты. Суммирование выходов квадратурных плеч дает модулированный сигнал, в котором пилотная компонента совпадает с произведением пилот-сигнала (т.е. константы) с четырехфазно манипулированной несущей.

4.6.4 Канал синхронизации

После вхождения в синхронизм с пилот-сигналом МС располагает информацией о фазе принимаемой несущей, временных границах чипов и периодов короткой ПСП (один период занимает $2^{15} / 1,2288 \cdot 10^6 = 26,666 \dots$ мс), тогда как для приема сообщений необходимо знать границы кадров (каждый кадр трафика занимает 20 мс). Соответствующие сведения наряду с другими, нужными для установления и поддержания соединения, МС получает по каналу синхронизации. В формате сообщения этого канала содержатся данные о точном времени в системе (SYS_TIME), значении циклического сдвига короткой ПСП данной БС (PILOT_PN), идентификаторы БС и ЦКПС (MSC), значение мощности сигнала в пилотном канале, пара-

метры "длинной" ПСП (LC_STATE), скорость передачи данных в канале персонального вызова (PRAT). Структура физического канала синхронизации поясняется рис. 4.16.

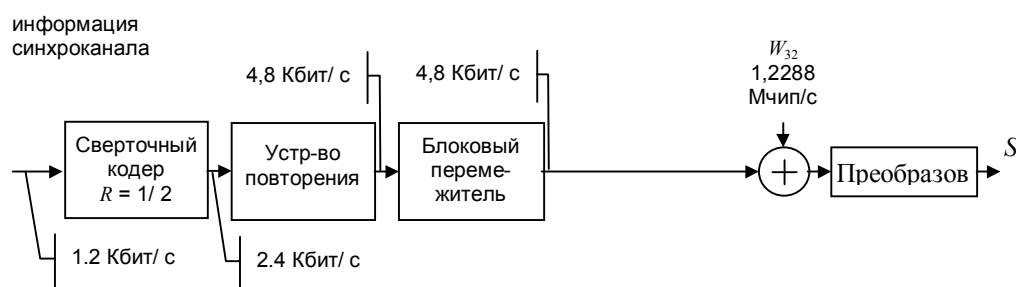


Рис. 4.16. Структурная схема канала синхронизации

Данные канала синхронизации, поступающие со скоростью 1200 бит/с, подаются на вход сверточного кодера. После осуществления операции кодирования (длина кодового ограничения – 9, скорость кода $1/2$) с выхода снимается двоичная последовательность со скоростью 2400 бит/с, поступающая далее на устройство повторения, после которого скорость потока данных удваивается до 4800 бит/с. Информация, передаваемая по каналу, структурирована в кадры длительностью в 26,666... мс. Каждый кадр совпадает с одним периодом короткой ПСП и содержит 32 бита исходных данных (128 символов с учетом сверточного кодирования и повторения). Как видно из рис. 4.16, следующей операцией является блочное перемежение в пределах кадра синхроканала, служащее для борьбы с пакетными ошибками. После блочного перемежителя поток данных подвергается прямому расширению спектра путем сложения по модулю два с присвоенной каналу синхронизации функцией Уолша w_{32} и преобразования булевых переменных в действительные $+1$ и -1 . Как можно видеть, каждому биту информационного потока с выхода перемежителя (скорость 4,8 кбит/с) сопоставляется $1,2288 \cdot 10^6 / 4,8 \cdot 10^3 = 256$ чипов, т.е. четыре периода последовательности Уолша. Иными словами, каждый бит кодированного потока в зависимости от значения заменяется четырьмя периодами прямой или инвертированной функции Уолша w_{32} .

Последующие операции в канале повторяют рассмотренные ранее: сигнал канала синхронизации объединяется с сигналами других каналов, поступая на вход S на рис. 4.14, после чего в КФМ модуляторе перемножается с комплексной короткой ПСП (скремблируется) и переносится на несущую.

Каждые три кадра (96 информационных бит) в канале синхронизации образуют один суперкадр длительности 80 мс, соответствующий четырем кадрам канала трафика (см. ниже). Сообщение, передаваемое по каналу синхронизации, может занимать более одного суперкадра, вследствие чего передаваемые данные подвергаются определенной структурной организации, называемой *капсулированием*. Полученная в результате подобного преобразования капсула состоит из самого сообщения синхроканала и добавочных бит, заполняющих оставшееся битовое пространство до начала следующего суперкадра. Необходимость подобной упаковки данных объясняется тем, что начало любого сообщения всегда должно совпадать с началом суперкадра.

Каждый кадр синхроканала начинается битом-заголовком, называемым *битом начала сообщения* (start-of-message – SOM). Значение этого бита, равное единице, указывает на начало сообщения, передаваемого по синхроканалу, а значение, равное нулю, свидетельствует о том, что текущий кадр содержит сообщение, начавшееся в некотором предыдущем кадре. В результате БС может передавать синхросообщение, занимающее несколько кадров подряд. Следует также отметить, что значение SOM-бита, равное 1, всегда должно совпадать с началом суперкадра, поскольку, как уже указывалось ранее, начало любого сообщения синхроканала всегда отвечает началу суперкадра. Ситуацию, соответствующую передаче по син-

хороканалу сообщения, занимающего два последовательных суперкадра, иллюстрирует рис. 4.17.



Рис. 4.17. Структура кадров канала синхронизации

Состав оборудования сетей стандарта CDMA IS-95 во многом сходен с составом оборудования сетей стандарта GSM и включает в себя ПС и БС, цифровые коммутаторы, центр управления и обслуживания, различные дополнительные системы и устройства, функциональное сопряжение элементов системы осуществляется с помощью ряда интерфейсов.

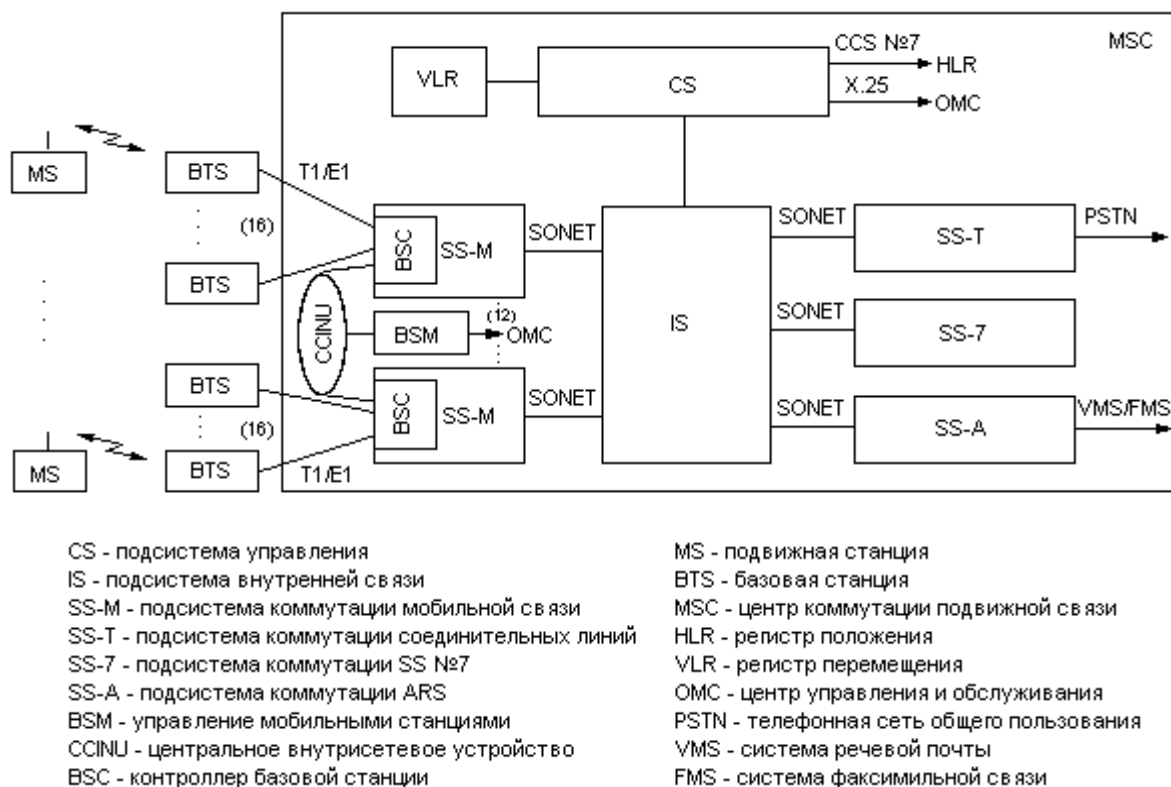


Рис. 4.18. Конфигурация системы стандарта IS-95

Конфигурация системы стандарта CDMA IS-95 представлена на рис. 4.18 (SONET – линия оптической связи).

каждая МС излучала сигнал минимальной мощности, которая достаточна для обеспечения приемлемого качества речи. За счет того, что все МС излучают сигналы необходимой для нормальной работы мощности, и не более, их взаимное влияние минимизируется, и абонентская емкость системы возрастает. МС должны обеспечивать регулирование выходной мощности в широком динамическом диапазоне - до 85 дБ.

При процедуре мягкой эстафетной передачи (переходе абонента из зоны обслуживания одной БС в зону другой) схема регулирования мощности несколько иная. МС принимает одновременно несколько команд управления мощностью от разных БС (обычно двух) и сравнивает их между собой. Если все команды указывают на необходимость увеличения мощности, то МС последовательно увеличивает свою мощность с шагом 1 дБ.

Регулирование мощности как в прямом, так и в обратном канале влияет на срок службы аккумуляторов МС. Средняя излучаемая мощность МС в CDMA меньше, чем в системах, использующих другие методы доступа. Это непосредственно связано с такими параметрами радиотелефона, как длительность непрерывного занятия канала и время нахождения в режиме ожидания.

4.7 Общая характеристика 3G – 4G поколений ССМС

Таблица 4.2 дает представление о характеристиках 1G – 4G поколений ССМС.
Таблица 4.2

Поколение	1G	2G	2.5G	3G	3.5G	4G
Начало разработок	1970	1980	1985	1990	<2000	2000
Реализация	1984	1991	1999	2002	2006-2007	2008
Сервисы	аналоговый стандарт, синхронная передача данных со скоростью до 9,6 кбит/с	цифровой стандарт, поддержка коротких сообщений (SMS)	большая ёмкость, пакетная передача данных	ещё большая ёмкость, скорости до 2 Мбит/с	увеличение скорости сетей третьего поколения	большая ёмкость, ориентированная на мультимедийные скорости передачи данных
Стандарты	AMPS, TACS, NMT и др.	TDMA, CDMA, GSM, PDC	GPRS, EDGE, 1xRTT	WCDMA, CDMA2000, UMTS	HSDPA	единица
Скорость передачи	1,9 кбит/с	14,4 кбит/с	384 кбит/с	2 Мбит/с	3-14 Мбит/с	1 Гбит/с
Сеть	PSTN*	PSTN*	PSTN*, сеть пакетной передачи данных	сеть пакетной передачи данных	сеть пакетной передачи данных	Интернет

*PSTN – Public Switched Telephone Network – ТФОП

В сотовой связи смена поколений выражена намного более ярко, чем, скажем, в индустрии персональных компьютеров или другой подобной техники. В мобильном мире все четко: 1G (первое поколение) — это аналоговая связь (стандарт NMT). 2G — поколение цифровой связи с коммутацией каналов (стандарты GSM и CDMA). Третье поколение — 3G (стандарт

UMTS) — предусматривает наряду с коммутацией каналов и пакетную передачу данных. Появились и промежуточные стандарты, призванные повысить скорость передачи сетей GSM; так, 2,5G определяет три основных дополнения к базовому стандарту:

1. **HSCSD** (*high-speed circuit-switched data*) – предоставление одному и тому же пользователю нескольких слотов в GSM кадре;
2. **GPRS** (*general packet radio service*) – предусматривает организацию специального канала пакетной передачи со скоростями до 115 кбит/с (или даже выше, но без помехоустойчивого кодирования);
3. **EDGE** (*enhanced data rates for global evolution*) подразумевает введение наряду с гауссовской МЧМ добавочного модуляционного формата: восьмеричной ФМ, утраивающей скорость передачи в рамках фиксированного частотного ресурса.

Совокупность всех мероприятий позволяет повысить скорость передачи до 384 кбит/с.

Субъектами разработки стандартов третьего поколения (3G) явились Международный Телекоммуникационный Союз **ITU** (*International Telecommunication Union*), Европейский институт телекоммуникационных стандартов **ETSI** (*European Telecommunication Standards Institute*), ассоциации радиопромышленников США (**TIA** – *Telecommunication Industry Association*) и Японии (**ARIB** – *Association of Radio Industries and Businesses*), органы, представляющие интересы телекоммуникационных сообществ Китая и Кореи, а также крупнейшие корпорации, определяющие конъюнктуру мирового рынка (Nokia, Eriksson, Qualcomm, Motorola, Lucent, Nortel и др.).

Стандарт 3G, разработанный ITU, носит название IMT 2000 (International Mobile Telecommunications 2000). Под этой аббревиатурой объединены пять стандартов, и только некоторые из них обеспечивают полное покрытие в различных диапазонах, поэтому фактически только они могут рассматриваться в качестве полноценных 3G решений.

Используются три основных стандарта 3G: UMTS (Universal Mobile Telecommunications Service), CDMA2000 и WCDMA (Wide CDMA). Все они настроены на пакетную передачу данных и, соответственно, на работу с цифровыми компьютерными сетями, включая Интернет.

4.7.1 Основные стандарты 3G

Для реализации систем третьего поколения разработаны рекомендации по глобальным унифицированным стандартам мобильной связи: обеспечение качества передачи речи, сравнимого с качеством передачи в проводных сетях связи; обеспечение безопасности, сравнимой с безопасностью в проводных сетях; обеспечение национального и международного роуминга; поддержка нескольких местных и международных операторов; эффективное использование спектра частот; пакетная и канальная коммутация; поддержка многоуровневых сетевых структур; взаимодействие с системами спутниковой связи; поэтапное наращивание скорости передачи данных вплоть до 2 Мбит/с.

Согласованные базовые требования, определяющие философию 3G, в укрупненной формулировке таковы:

- Скорость передачи данных до 2 Мбит/с в пределах полосы не более 5 МГц;
- Возможность варьирования скорости в широком диапазоне в зависимости от характера передаваемых данных;
- Возможность мультимедийной передачи в рамках одного контакта данных разнообразного содержания и разного уровня требований к качеству (речь, мультимедиа, пакеты и т.п.);
- Сосуществование систем 2G и 3G и возможность эстафетной передачи между ними с целью расширения зон покрытия и выравнивания трафика;
- Поддержка асимметричного режима работы, когда линия "вниз" имеет значительно более напряженный трафик (например, при поиске информации в Интернете) по сравнению с линией "вверх";
- Возможность реализации дуплексной передачи в вариантах как частотного, так и вре-

менного разноса линий "вниз" и "вверх";

- Максимальная гибкость сетевого оборудования и возможность построения его на основе "набора инструментов" (*toolbox*);
- Доступность мобилизации наиболее современных резервов улучшения качества связи (адаптация к условиям распространения, пространственная селекция терминала – *smart* антенны, многопользовательские – *multiuser* – приемники и т.д.).

К середине 2002 г. лидирующие позиции закрепились за двумя проектами, общей платформой которых является технология кодового разделения (**CDMA**). Первый из них – **UMTS** (*Universal Mobile Telecommunication System*) является продуктом кооперации Евросоюза (в лице ETSI) и Японии (ARIB) и имеет действенную поддержку со стороны Кореи, Китая, а также многих корпораций Европы и США.

Хотя радиоинтерфейс UMTS, получивший название UTRAN (*Universal Terrestrial Radio Access Network*), основан на кодовом разделении и практически не имеет ничего общего с физическим уровнем GSM, на сетевом уровне предлагаемый стандарт целиком опирается на "готовую" инфраструктуру GSM.

Второй проект, известный под именем **cdma2000**, является максимально преемственным по отношению к системе **cdmaOne (IS-95)** и настойчиво продвигается компаниями Qualcomm, Lucent и Motorola – основными поставщиками и операторами действующих CDMA систем.

CDMA2000 является стандартом 3G в эволюционном развитии сетей cdmaOne. При сохранении основных принципов, заложенных версией IS-95A, технология стандарта CDMA непрерывно развивается.

При построении системы мобильной связи на основе технологии CDMA2000 1X первая фаза обеспечивает передачу данных со скоростью до 153 кбит/с, что позволяет предоставлять услуги голосовой связи, передачу коротких сообщений, работу с электронной почтой, Интернетом, базами данных, передачу данных и неподвижных изображений. Переход к следующей фазе CDMA2000 1xEV-DO происходит при использовании той же полосы частот 1,23 МГц, скорость передачи — до 2,4 Мбит/с в прямом канале и до 153 кбит/с в обратном, что делает эту систему связи отвечающей требованиям 3G и даёт возможность предоставлять самый широкий спектр услуг, вплоть до передачи видео в режиме реального времени. Следующей фазой развития стандарта в направлении увеличения сетевой ёмкости и передачи данных является 1xEV-DO Rev A: передача данных со скоростью до 3,1 Мбит/с по направлению к абоненту и до 1,8 Мбит/с — от абонента. Операторы смогут предоставлять те же услуги, что и на базе Rev. 0, а, кроме того, передавать голос, данные и осуществлять широко вещание по IP сетям. Следующая фаза — 1xEV-DO Rev B, — разрабатывается с целью достигнуть более высоких скоростей на одном частотном канале: 4,9 Мбит/с к абоненту и 2,4 Мбит/с от абонента. К тому же будет обеспечиваться возможность объединения нескольких частотных каналов для увеличения скорости. Например, объединение 15-ти частотных каналов (максимально возможное количество) позволит достигать скоростей 73,5 Мбит/с к абоненту и 27 Мбит/с от абонента. Применение таких сетей — улучшенная работа чувствительных к временным задержкам приложений типа VoIP, Push to Talk, видеотелефония, сетевые игры и т.п.

Основными компонентами коммерческого успеха системы CDMA2000 являются более широкая зона обслуживания, высокое качество речи (практически эквивалентное проводным системам), гибкость и дешевизна внедрения новых услуг, высокая помехозащищённость, устойчивость канала связи от перехвата и прослушивания.

WCDMA (англ. Wideband Code Division Multiple Access — широкополосный CDMA) — технология радиоинтерфейса, избранная большинством операторов сотовой связи для обеспечения широкополосного радиодоступа с целью поддержки услуг 3G.

Технология оптимизирована для предоставления высокоскоростных мультимедийных услуг типа видео, доступа в Интернет и видеоконференций; обеспечивает скорости доступа вплоть до 2 Мбит/с на коротких расстояниях и 384 Кбит/с на больших с полной мобильно-

стью. Такие величины скорости передачи данных требуют широкую полосу частот, поэтому ширина полосы WCDMA составляет 5 МГц.

Технология может быть добавлена к существующим сетям GSM и PDC, что делает стандарт WCDMA наиболее перспективным с точки зрения использования сетевых ресурсов и глобальной совместимости.

Использование выделенных ресурсов различными технологиями показано на рис. 4.20, который дает также наглядное представление об их отличиях.

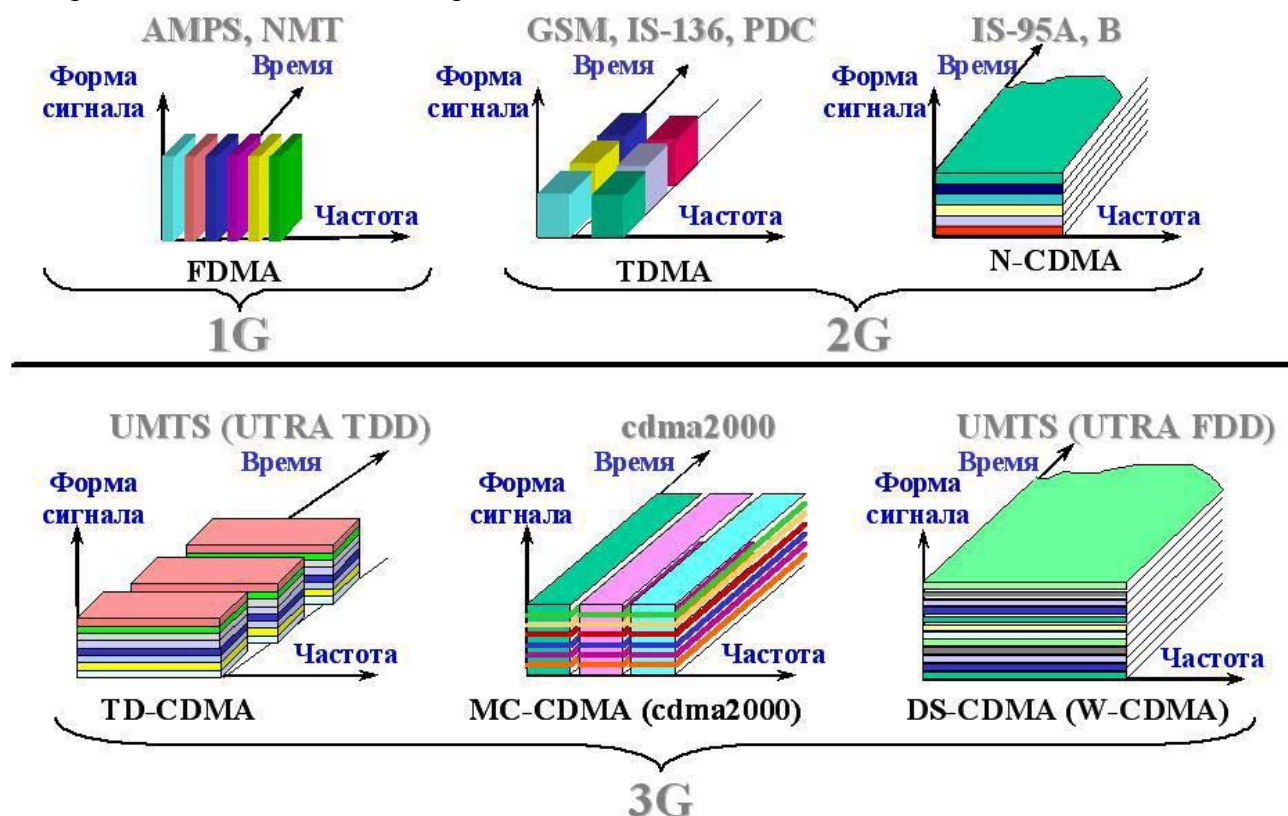


Рис. 4.20. Множественный доступ в мобильных сотовых сетях 1G – 3G.

Главным отличием 3G от сетей второго поколения является индивидуализация, то есть присвоение каждому абоненту IP-адреса, подобно Интернету. Еще один плюс – абоненты могут находиться в сети постоянно, не беспокоясь о материальных средствах, так как оплата насчитывается не за время, а за трафик.

Для завершения работы над спецификацией стандарта создано организационное ядро **3GPP** (*3G Partnership Project*), наименование которого часто используется как синоним названия самого проекта UMTS.

Ключевой технологией для UMTS является WCDMA.

Спецификация 3GPP определяет UMTS как широкополосную (*spread spectrum*) систему с кодовым разделением (CDMA) и прямым (*direct sequence*) расширением спектра, жестко оговаривая структуры **расширяющих, канализирующих и скремблирующих** кодов, а также устанавливая единую и неизменную частоту следования чипов названных кодов в 3,84 Мчип/с ($3,84 \cdot 10^6$ чипов в секунду).

Под **скремблированием** понимается процесс наложения на коды символов открытого текста кодов случайной последовательности чисел, которую называют также гаммой (по названию буквы γ греческого алфавита, используемой в математических формулах для обозначения случайного процесса). Гаммирование относится к поточным методам шифрования, когда следующие друг за другом символы открытого текста последовательно превращаются в символы шифрограммы, что повышает скорость преобразования.

В качестве гаммы широко используются *псевдослучайные последовательности* (ПСП). Коды символов криптограммы Y_i есть поразрядная сумма по модулю 2 кодов X_i и Z_i . Расшифровка производится по правилу: $\hat{X} = Y + Z = X + Z + Z = X$, так что на приемной стороне должна генерироваться такая же ПСП и с той же фазой. Стойкость скремблирования определяется параметрами ПСП, главным из которых является период последовательности. На практике используются генераторы ПСП с периодом в несколько недель.

4.7.2 Промежуточный стандарт 3,5 G

Начиная с 2006 года, в сетях UMTS повсеместно распространяется технология высокоскоростной пакетной передачи данных от базовой станции к мобильному терминалу HSDPA, которую принято относить к сетям поколения «3,5G». К началу 2008 года HSDPA поддерживала скорость передачи данных в режиме «от базовой станции к мобильному терминалу» до 7,2 Мбит/сек. Также ведутся разработки по повышению скорости передачи данных в режиме от мобильного терминала к базовой станции HSUPA. В долгосрочной перспективе, согласно проектам 3GPP, планируется эволюция UMTS в сети четвертого поколения 4G, позволяющие базовым станциям передавать и принимать информацию на скоростях 100 Мбит/сек и 50 Мбит/сек соответственно, благодаря усовершенствованному использованию воздушной среды посредством мультиплексирования с ортогональным частотным разделением сигналов OFDM.

UMTS позволяет пользователям проводить сеансы видеоконференций посредством мобильного терминала, однако опыт работы операторов связи Японии и некоторых других стран показал невысокий интерес абонентов к данной услуге. Гораздо более перспективным представляется развитие сервисов, предлагающих загрузку музыкального и видео контента: высокий спрос на услуги такого рода был продемонстрирован в сетях 2,5G.

HSDPA (англ. High-Speed Downlink Packet Access — высокоскоростная пакетная передача данных от базовой станции к мобильному телефону) — стандарт мобильной связи, рассматривается специалистами как один из переходных этапов миграции к технологиям мобильной связи четвертого поколения (4G). Максимальная теоретическая скорость передачи данных по стандарту составляет 14,4 Мбит/сек, практически же достижимая скорость в существующих сетях обычно не превышает 6 Мбит/сек.

Стандартом предусмотрено 20 категорий с различными значениями максимальной скорости передачи данных в пакетном режиме от базовой станции к мобильному терминалу. Каждая категория характеризуется максимальным числом одновременно используемых кодов (до 15), и типом модуляции в радиоканале (QPSK/16 QAM).

Категории 1-5 используют до 5 кодов и имеют максимальную скорость 1,2-3,6 Мб/с, категория 1 — до 14 Мб/с.

HSURA — аналогично HSDPA, технология высокоскоростной пакетной передачи данных по направлению «вверх» (англ. High-Speed Uplink Packet Access) представляет собой стандарт мобильной связи, позволяющий ускорить передачу данных от W-CDMA устройств конечного пользователя до базовой станции за счёт применения более совершенных методов модуляции.

Теоретически стандарт HSUPA рассчитан на максимальную скорость передачи данных по направлению «вверх» до 5,7 Мбит/с, позволяя, таким образом, использовать приложения третьего поколения, требующие обработки огромных потоков данных от мобильного устройства к базовой станции, например, видеоконференцсвязь.

Описание технологии планируется ввести в качестве спецификации 6-й версии стандарта 3GPP Release 6; процесс стандартизации технологии приближается к завершению.

4.7.3 Широкополосные технологии доступа к интернет-данным

К связи четвертого поколения (4G), как правило, относят технологии, которые позволяют передавать данные в сотовых сетях со скоростью выше 100 Мбит/сек.

Технология LTE (Long-Term Evolution) – это основное направление эволюции сетей сотовой связи третьего поколения (3G). В январе 2008 года международное объединение Third Generation Partnership Project (3GPP), разрабатывающее перспективные стандарты мобильной связи, утвердило LTE в качестве следующего после UMTS стандарта широкополосной сети мобильной связи.

Сети 4G на основе стандарта LTE способны работать практически по всей ширине спектра частот от 700 МГц до 2,7 ГГц.

LTE обеспечивает теоретическую пиковую скорость передачи данных до 326,4 Мбит/с от базовой станции к пользователю и до 172,8 Мбит/с в обратном направлении.

Технология Long Term Evolution, как ожидается, приведет к появлению качественно новых мобильных сервисов: пользователи смогут в режиме реального времени получать видео высокого качества, работать с интерактивными службами и пр.

В широком понимании к 4G относят также технологии беспроводной передачи интернет-данных Wi-Fi (скоростные варианты этого стандарта) и WiMAX (в теории скорость может превышать 1 Гбит/сек).

Wi-Fi (Wireless Fidelity) – современная беспроводная технология соединения компьютеров в сеть или подключения их к Интернету. Wi-Fi разработан консорциумом Wi-Fi Alliance. Стандарт Wi-Fi позволяет предоставить высокоскоростной доступ ко всем ресурсам сети Интернет (электронная почта, Интернет-серфинг, ICQ и т.д.) с ноутбука, смартфона или КПК в зоне покрытия сети Wi-Fi. Технология обеспечивает одновременную работу в сети нескольких десятков активных пользователей, скорость передачи информации для конечного абонента может достигать 54 Мбит/с.

Стандарт WiMAX (аббревиатура от Worldwide Interoperability for Microwave Access) – технология широкополосной беспроводной связи, которая в отличие от других технологий радиодоступа, обеспечивает высокоскоростные соединения на больших расстояниях даже при отсутствии прямой видимости объекта, на отраженном сигнале.

Основана на стандарте IEEE 802.16, который также называют Wireless MAN. Название «WiMAX» было создано WiMAX Forum — организацией, которая была основана в июне 2001 года с целью продвижения и развития технологии WiMAX. Форум описывает WiMAX как «основанную на стандарте технологию, предоставляющую высокоскоростной беспроводной доступ к сети, альтернативный выделенным линиям и DSL».

WiMAX позволяет осуществлять доступ в Интернет на высоких скоростях, с гораздо большим покрытием, чем у Wi-Fi сетей. Это позволяет использовать технологию в качестве «магистральных каналов», продолжением которых выступают традиционные DSL- и выделенные линии, а также локальные сети. В результате подобный подход позволяет создавать масштабируемые высокоскоростные сети в рамках целых городов.

Фиксированный и мобильный вариант WiMAX

Набор преимуществ присущ всему семейству WiMAX, однако его версии существенно отличаются друг от друга. Разработчики стандарта искали оптимальные решения как для фиксированного, так и для мобильного применения, но совместить все требования в рамках одного стандарта не удалось. Хотя ряд базовых требований совпадает, нацеленность технологий на разные рыночные ниши привела к созданию двух отдельных версий стандарта (вернее, их можно считать двумя разными стандартами). Каждая из спецификаций WiMAX определяет свои рабочие диапазоны частот, ширину полосы пропускания, мощность излучения, методы передачи и доступа, способы кодирования и модуляции сигнала, принципы по-

вторного использования радиочастот и прочие показатели. А потому WiMAX-системы, основанные на версиях стандарта IEEE 802.16 e и d, практически несовместимы. Краткие характеристики каждой из версий приведены ниже.

802.16-2004 (известен также как 802.16d и фиксированный WiMAX). Спецификация утверждена в 2004 году. Используется ортогональное частотное мультиплексирование (OFDM), поддерживается фиксированный доступ в зонах с наличием либо отсутствием прямой видимости. Пользовательские устройства представляют собой стационарные модемы для установки вне и внутри помещений, а также PCMCIA-карты для ноутбуков. В большинстве стран под эту технологию отведены диапазоны 3,5 и 5 ГГц. По сведениям WiMAX Forum, насчитывается уже порядка 175 внедрений фиксированной версии. Многие аналитики видят в ней конкурирующую или взаимодополняющую технологию проводного широкополосного доступа DSL.

802.16-2005 (известен также как 802.16e и мобильный WiMAX). Спецификация утверждена в 2005 году. Это — новый виток развития технологии фиксированного доступа (802.16d). Оптимизированная для поддержки мобильных пользователей версия поддерживает ряд специфических функций, таких как хэндовер(англ.), idle mode и роуминг. Применяется масштабируемый OFDM-доступ (SOFDMA), возможна работа при наличии либо отсутствии прямой видимости. Планируемые частотные диапазоны для сетей Mobile WiMAX таковы: 2,3-2,5; 2,5-2,7; 3,4-3,8 ГГц. В мире реализованы несколько пилотных проектов, в том числе первым в России свою сеть развернул «Скартел». Конкурентами 802.16e являются все мобильные технологии третьего поколения (например, EV-DO, HSDPA).

Основное различие двух технологий состоит в том, что фиксированный WiMAX позволяет обслуживать только «статичных» абонентов, а мобильный ориентирован на работу с пользователями, передвигающимися со скоростью до 120 км/ч. Мобильность означает наличие функций роуминга и «бесшовного» переключения между базовыми станциями при передвижении абонента (как происходит в сетях сотовой связи). В частном случае мобильный WiMAX может применяться и для обслуживания фиксированных пользователей.

Пользовательское оборудование

Оборудование для использования сетей WiMAX поставляется несколькими производителями и может быть установлено как в помещении (устройства размером с обычный DSL модем), так и вне него (устройства размером с флэш-накопитель). Следует заметить что оборудование, рассчитанное на размещение внутри помещений и не требующее профессиональных навыков при установке, конечно, более удобно, однако способно работать на значительно меньших расстояниях от базовой станции, чем профессионально установленные внешние устройства. Поэтому оборудование, установленное внутри помещений требует намного больших инвестиций в развитие инфраструктуры сети, так как подразумевает использование намного большего числа точек доступа.

С изобретением мобильного WiMAX все больший акцент делается на разработке мобильных устройств. В том числе специальных телефонных трубок (похожи на обычный мобильный смартфон), и компьютерной периферии (USB радио модулей и PC card).

Принцип работы Основные понятия

В общем виде WiMAX сети состоят из следующих основных частей: базовых и абонентских станций, а также оборудования, связывающего базовые станции между собой, с поставщиком сервисов и с Интернетом.

Для соединения базовой станции с абонентской используется высокочастотный диапазон радиоволн от 1,5 до 11 ГГц. В идеальных условиях скорость обмена данными может дос-

тигать 70 Мбит/с, при этом не требуется обеспечения прямой видимости между базовой станцией и приемником.

Как уже говорилось выше, WiMAX применяется как для решения проблемы «последней мили», так и для предоставления доступа в сеть офисным и районным сетям[1].

Между базовыми станциями устанавливаются соединения (прямой видимости), использующие диапазон частот от 10 до 66 ГГц, скорость обмена данными может достигать 120 Мбит/с. При этом, по крайней мере одна базовая станция подключается к сети провайдера с использованием классических проводных соединений. Однако, чем большее число БС подключено к сетям провайдера, тем выше скорость передачи данных и надёжность сети в целом.

Структура сетей семейства стандартов IEEE 802.16 схожа с традиционными GSM сетями (базовые станции действуют на расстояниях до десятков километров, для их установки не обязательно строить вышки — допускается установка на крышах домов при соблюдении условия прямой видимости между станциями)[1].

Сравнение Wi-Fi и WiMAX

Сопоставления WiMAX и Wi-Fi далеко не редкость, возможно, потому, что эти термины созвучны, название стандартов, на которых основаны эти технологии, похожи (стандарты IEEE, оба начинаются с «802.»), а также обе технологии используют беспроводное соединение и используются для подключения к Интернету (каналу обмена данными). Но несмотря на это, эти технологии направлены на решение совершенно различных задач.

Сравнительная таблица стандартов беспроводной связи

Технология	Стандарт	Использование	Пропускная способность	Радиус действия	Частоты
Wi-Fi	802.11a	WLAN	до 54 Мбит/с	до 100 метров	5,0 ГГц
Wi-Fi	802.11b	WLAN	до 11 Мбит/с	до 100 метров	2,4 ГГц
Wi-Fi	802.11g	WLAN	до 108 Мбит/с	до 100 метров	2,4 ГГц
Wi-Fi	802.11n	WLAN	до 300 Мбит/с (в перспективе до 450, а затем до 600 Мбит/с)	до 100 метров	2,4 — 2,5 или 5,0 ГГц
WiMax	802.16d	WMAN	до 75 Мбит/с	6-10 км	1,5-11 ГГц
WiMax	802.16e	Mobile WMAN	до 30 Мбит/с	1-5 км	2-6 ГГц
WiMax	802.16m	WMAN, Mobile WMAN	до 1 Гбит/с (WMAN), до 100 Мбит/с (Mobile WMAN)	н/д (стандарт в разработке)	н/д (стандарт в разработке)
Bluetooth v. 1.1.	802.15.1	WPAN	до 1 Мбит/с	до 10 метров	2,4 ГГц
Bluetooth v. 1.3.	802.15.3	WPAN	от 11 до 55 Мбит/с	до 100 метров	2,4 ГГц

WiMAX это система дальнего действия, покрывающая километры пространства, которая обычно использует лицензированные спектры частот (хотя возможно и использование нелицензированных частот) для предоставления соединения с Интернетом типа точка-точка провайдером конечному пользователю. Разные стандарты семейства 802.16 обеспечивают разные виды доступа, от мобильного (схож с передачей данных с мобильных телефонов) до фиксированного (альтернатива проводному доступу, при котором беспроводное оборудование пользователя привязано к местоположению).

Wi-Fi это система более короткого действия, обычно покрывающая десятки метров, которая использует нелицензированные диапазоны частот для обеспечения доступа к сети. Обычно Wi-Fi используется пользователями для доступа к их собственной локальной сети, которая может быть и не подключена к Интернету. Если WiMAX можно сравнить с мобильной связью, то Wi-Fi скорее похож на стационарный беспроводной телефон.

Bluetooth – технология беспроводной ближней коротковолновой радиосвязи (до 30 м), позволяющая объединять устройства разных типов для передачи речи и данных. Её разработкой и развитием занимается ассоциация Bluetooth SIG . Стандарт получил обозначение IEEE 802.15 . Он определяет работу на частоте 2,4 ГГц, со скоростями передачи 722-784 Кбит/с и расстояниями до 10 м. Свое название технология получила в честь датского короля Гаральда Синий Зуб, правившего Данией и Норвегией в X веке

WiMAX и Wi-Fi имеют совершенно разный механизм Quality of Service (QoS). WiMAX использует механизм, основанный на установлении соединения между базовой станцией и устройством пользователя. Каждое соединение основано на специальном алгоритме планирования, который может гарантировать параметр QoS для каждого соединения. Wi-Fi, в свою очередь, использует механизм QoS подобный тому, что используется в Ethernet, при котором пакеты получают различный приоритет. Такой подход не гарантирует одинаковый QoS для каждого соединения.

Из-за дешевизны и простоты установки, Wi-Fi часто используется для предоставления клиентам быстрого доступа в Интернет различными организациями. Например, в некоторых кафе, отелях, вокзалах и аэропортах можно обнаружить бесплатную точку доступа Wi-Fi.

Стандарт WiMAX был разработан корпорацией Intel, крупнейшим мировым производителем микрочипов. Соответственно, WiMAX-чипами будут в первую очередь комплектовать ноутбуки. Скорей всего, со временем WiMAX вытеснит Wi-Fi, так как Wi-Fi действует в радиусе нескольких метров от точки доступа, у мобильного WiMAX покрытие существенно больше. И кроме того, он позволяет абоненту, если тот перемещается со скоростью до 120 км/ч, переключаться между станциями.

Список литературы

1. Склад Б. Цифровая связь. Теоретические основы и практическое применение. Изд. 2-е, испр.: Пер. с англ. – М.: Издательский дом «Вильямс», 2003. – 1104 с.
2. Ипатов В. Широкополосные системы и кодовое разделение сигналов. Принципы и приложения. Москва: Техносфера, 2007. – 488 с.
3. Прокис Дж. Цифровая связь. Пер. с англ. / Под ред. Д.Д. Кловского – М.: Радио и связь, 2000. – 800 с.
4. Пенин П.И. Системы передачи цифровой информации: Учебное пособие для вузов.-М.: Сов. радио, 1983. – 368с.
5. Системы мобильной связи: Учебное пособие для вузов/В. П. Ипатов, В. И. Орлов, И. М. Самойлов, В. Н. Смирнов; под. ред. В. П. Ипатова. -М.: Горячая линия – Телеком, 2003, -272с.
6. Градштейн И.С., Рыжик И.М.. Таблицы интегралов, сумм, рядов и произведений. Москва: - Наука, 1971. – 1109 с.
7. Турин Дж.Л. Введение в широкополосные методы борьбы с многолучевостью распространения радиосигналов и их применение в городских системах цифровой связи. ТИИЭР, т. 68, № 2, март 1980. С. 30 – 60.
8. Файфер У.С., Бруно Ф.Дж. Недорогая приемопередающая установка пакетной радиосвязи. ТИИЭР, т. 75, № 1, январь 1987. С. 41 – 53.
9. Невдяев Л.М. Мобильная связь 3-го поколения. М.: «Связь и бизнес», МЦНТИ, 2000 – 208 с.

