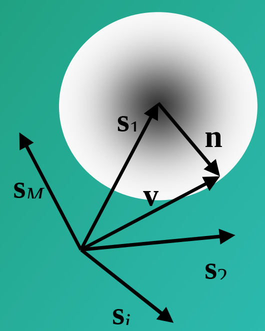
Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования

Новгородский государственный университет
имени Ярослава Мудрого

Рассветалов Л.А.

СТАТИСТИЧЕСКАЯ ТЕОРИЯ СВЯЗИ



Великий Новгород
2014 г.

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

Новгородский государственный университет
имени Ярослава Мудрого

Рассветалов Л.А.

СТАТИСТИЧЕСКАЯ ТЕОРИЯ СВЯЗИ

УЧЕБНОЕ ПОСОБИЕ

Великий Новгород
2014 г.

ББК

УДК 621.396; 681.3.07

Рассветалов Л.А. Статистическая теория связи: Учебное пособие / ФГБОУ «Новгородский государственный университет им. Ярослава Мудрого», Великий Новгород, 2014 г. - 112 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,.

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.

Содержание

Предисловие

Пособие состоит из трех 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 [1, 2].

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 [5]. В конце этой главы помещено краткое описание установки пакетной радиосвязи [6], в которой применен остроумный способ создания опорного напряжения для детектирования широкополосных сигналов с абсолютной фазовой манипуляцией в многолучевом канале подвижной радиосвязи.

1 Структурная схема цифровой связи

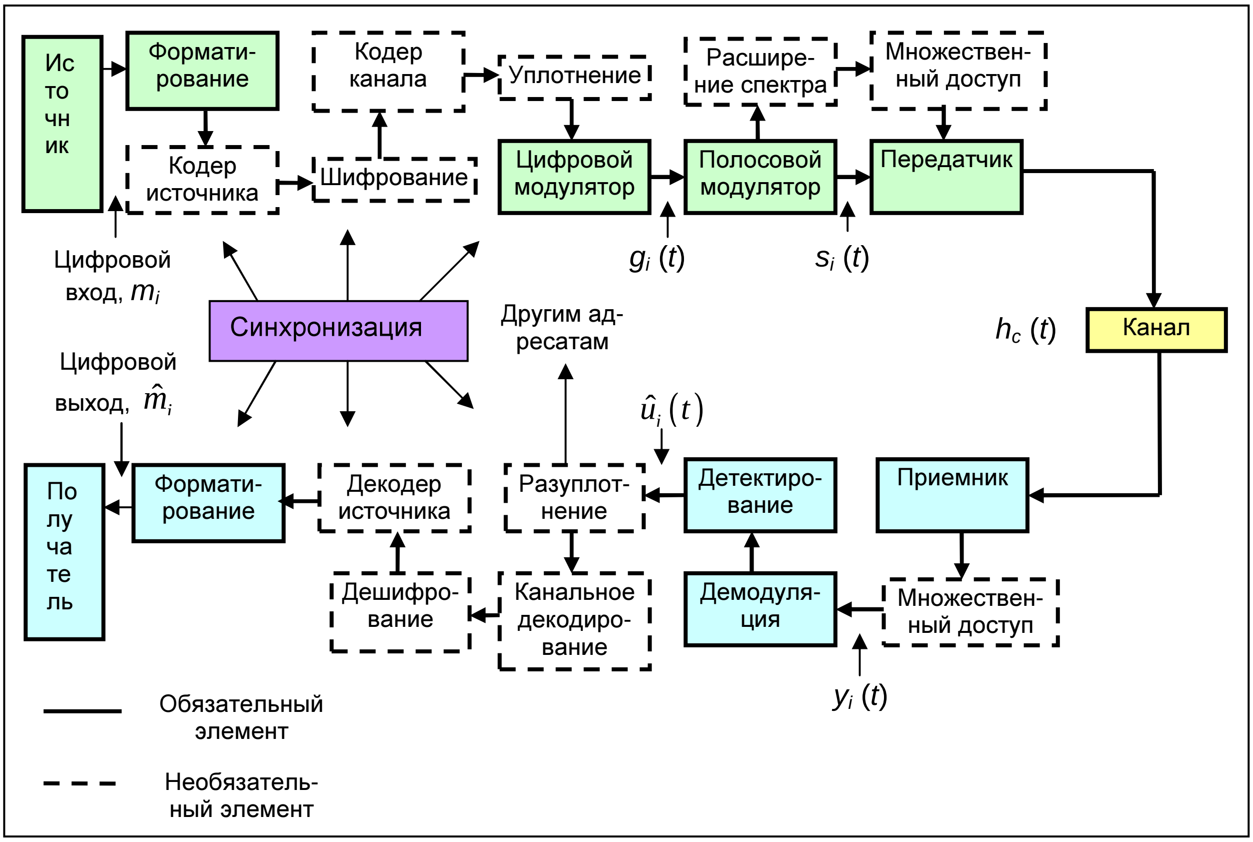


Рис.1.1. Основные элементы цифровой системы связи [1]

На рис.1.1 представлена функциональная схема и основные элементы цифровой системы связи (ЦСС). Выход источника может быть либо аналоговым сигналом (звуковым или видео), либо цифровым, дискретным во времени и имеющим конечное число выходных значений. В системах цифровой связи сообщения источника преобразуются в двоичные символы (биты); после этого биты группируются *в цифровые сообщения* или *символы сообщений*. Каждый такой символ (m_i, где i = **1, …,** $\boldsymbol{M}$**)** можно рассматривать как элемент конечного алфавита, содержащего $\boldsymbol{M}$ элементов. Следовательно, для $\boldsymbol{M}$ = 2 символ сообщения m_i является бинарным (т.е. состоит из одного бита).

В верхних блоках осуществляется преобразование сигнала на пути от источника к передатчику - форматирование, кодирование источника, шифрование, канальное кодирование, уплотнение, импульсная модуляция, полосовая модуляция, расширение спектра и множественный доступ. Нижние блоки показывают преобразования сигналов на пути от приемника к получателю сообщения, и они противоположны верхним блокам.

Блоки *модулятор/демодулятор* образуют *модем*, а *кодер/декодер – кодек*. Термины «модем» и «кодек» могут включать в себя несколько этапов обработки сигналов, как это показано на рис. 1.1. Поясним введенные выше термины и операции преобразования сигналов.

Форматирование преобразовывает информацию источника в биты, т.е. осуществляет аналого-цифровое преобразование, если сигнал источника аналоговый. На рис. 1.1 эта операция не показана; предполагается, что она осуществляется в самом источнике, либо в кодере источника.

Кодирование источника осуществляется, как правило, с целью уменьшения избыточности его сообщений; в этом случае оно называется *статистическим* или *эффективным*.

Шифрование осуществляется для увеличения криптостойкости передаваемых сообщений.

Канальное 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. Кодированные сигналы разных источников в блоке *уплотнения* образуют *групповой сигнал*.

Помехоустойчивое кодирование может осуществляться как при канальном кодировании, так и над групповым сигналом. Специфические вопросы многоканальной передачи здесь рассматриваться не будут; дальнейшее описание предполагает наличие одного источника (канала).

Импульсная (*цифровая*) *модуляция* — это еще один необходимый этап, поскольку каждый символ, который требуется передать, вначале нужно преобразовать из двоичного представления в *видеосигнал (модулированный сигнал).* Блок цифровой 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 недвоичным символам результирующий сигнал именуется М-арным (М-ичным) импульсно-модулированным. После цифровой модуляции каждый символ сообщения или канальный символ принимает форму видеосигнала $g_i(t)$, где i = 1, ..., M.

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, ..., 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 $s_i(t)$:

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

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. На этом этапе обработки обычно выполняется *выравнивание* – 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*. 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 ($i = 1, ..., M$) 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 видеоимпульсов $g_1(t), g_2(t)...., g_M(t)$. 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 $s_1(t), s_2(t), ..., s_M(t)$. 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 *потоком 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 *М-арной* (*М-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 "нуль". 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 Ђ	144 ђ	160	176 °	192 А	208 Р	224 а	240 р
129 Ѓ	145 ‘	161 Ў	177 ±	193 Б	209 С	225 б	241 с
130 ‚	146 ’	162 ў	178 І	194 В	210 Т	226 в	242 т
131 ѓ	147 “	163 Ј	179 і	195 Г	211 У	227 г	243 у
132 „	148 ”	164 ¤	180 ґ	196 Д	212 Ф	228 д	244 ф
133 …	149 •	165 Ґ	181 µ	197 Е	213 Х	229 е	245 х
134 †	150 –	166 ¦	182 ¶	198 Ж	214 Ц	230 ж	246 ц
135 ‡	151 —	167 §	183 ·	199 З	215 Ч	231 з	247 ч
136 ˚	152	168 Ё	184 ё	200 И	216 Ш	232 и	248 ш
137 ‰	153 ™	169 ©	185 №	201 Й	217 Щ	233 й	249 щ
138 Љ	154 љ	170 Є	186 є	202 К	218 Ъ	234 к	250 ъ
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 Ѕ	205 Н	221 Э	237 н	253 э
142 Ћ	158 ћ	174 ®	190 ѕ	206 О	222 Ю	238 о	254 ю
143 Џ	159 џ	175 Ї	191 ї	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 $s_1(t)$, где i = 1, ..., 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!»

Знаковое кодирование Д У М А Й !

Кодировка Windows 1251 110001001101001111001100110000001100100110000100

8-ричные цифры 6 1 1 5 1 7 1 4 6 0 1 4 4 6 0 4

8-ричные сигналы $s_1(t)$ $s_2(t)$ $s_3(t)$ $s_4(t)$ $s_5(t)$ $s_6(t)$ $s_7(t)$ $s_8(t)$ $s_9(t)$ $s_{10}(t)$ $s_{11}(t)$ $s_{12}(t)$ $s_{13}(t)$ $s_{14}(t)$ $s_{15}(t)$ $s_{16}(t)$

а)

Знаковое кодирование Д У М А Й !

Кодировка Windows 1251 110001001101001111001100110000001100100110000100

32-ричные числа 24 19 9 28 25 16 6 9 16 16

32-ричные сигналы $s_1(t)$ $s_2(t)$ $s_3(t)$ $s_4(t)$ $s_5(t)$ $s_6(t)$ $s_7(t)$ $s_8(t)$ $s_9(t)$ $s_{10}(t)$

б)

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 $i=1, \ldots, 32$, 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

$$\beta_E = E_b / N_0, \tag{1.2}$$

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

$$\beta_W = W/R, \tag{1.3}$$

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)

1.3 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. 1.3).

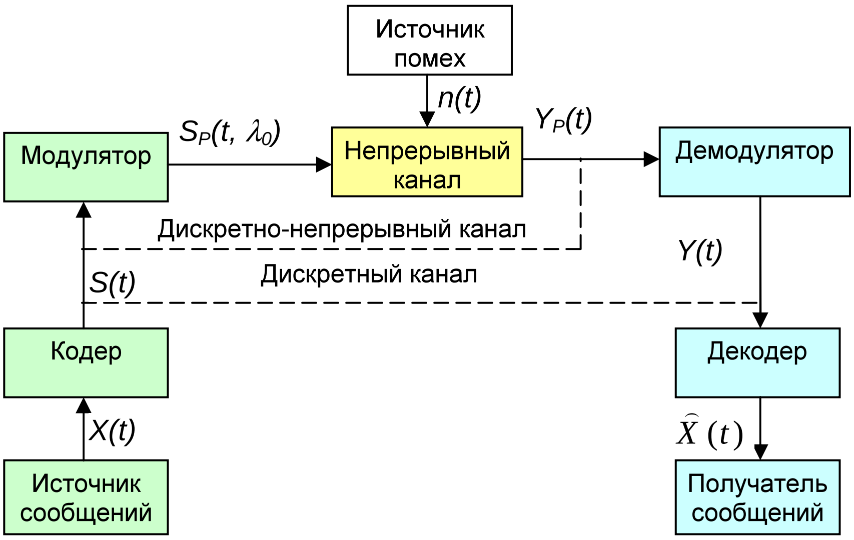


Рис. 1.3. Обобщенная схема цифровой системы передачи.

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 $n(t)$:

$$Y_P(t) = h_c * S_P(t, \lambda_0) + n(t) = S_P(t, \lambda) + n(t) \qquad (1.4)$$

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:

$$S_P(t, \lambda_0) = A(t)cos[\omega_0 t+\varphi(t)], \tag{1.5}$$

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 $\tau_з$, доплеровским смещением частоты Ω, случайной начальной фазой θ и имеет вид

$$S_P(t, \lambda) = \alpha(t)A(t - \tau_з)cos[(\omega_0 +\Omega) t +\varphi(t - \tau_з) - \theta]. \tag{1.6}$$

Т.о., совокупность параметров принятого полезного сигнала

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

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 несущественны, 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

$$Y_p(t) = S_p (t, \lambda_0) + n(t). \tag{1.7}$$

2. **Гауссовский канал с неизвестной фазой** сигнала определяется параметрами Ω, $\tau_з$, $\alpha(t) = \alpha$, которые постоянны и известны. Начальная фаза неизвестна и обычно считается равномерно распределенной в интервале (0, 2π). 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

$$S_P(t, \lambda) = \alpha(t)A(t - \tau_з)cos[(\omega_0 + \Omega)t + \varphi(t) - \theta]. \quad (1.8)$$

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 (1.8) 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. 1.3 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 осуществляется модулятором; 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 $Y = \{y_1, y_2, y_3, \ldots, y_m\}$ в виде суммы переданной последовательности $S=\{s_1, s_2, s_3, \ldots, s_m\}$ и комбинации помехи (вектора ошибки) $E = \{e_1, e_2, e_3, \ldots, e_m\}$

$$Y = S \oplus E, \qquad (1.9)$$

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.

Стационарных 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. 1.4.

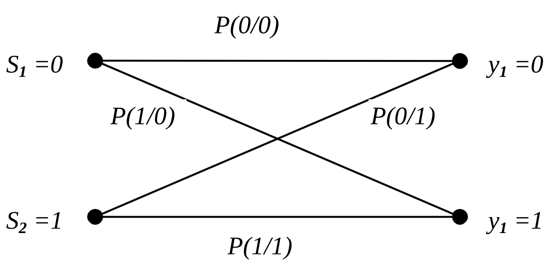


Рис. 1.4. Граф двоичного канала

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 (1.9), 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:

$$P_n(t) = C^t_n\, P^*_n(t) = n!p^t\,(1-p)^{n-t}/t!(n-t)! \qquad (1.10)$$

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

$$P(y_i/s_i, s_{i-1}, \dots, s_{i-N}) = P(y_i/s_i, s_{i-1}, \dots, s_{i-N-j}), \quad j \geq 1 \qquad (1.11)$$

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

$$P(y_i, C_i/s_i, C_{i-1}) = P(y_i/s_i, C_{i-1})\,P(C_i/C_{i-1}) \qquad (1.12)$$

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 $P^{(2)}_e$=0,4).

Канал можно описать матрицей переходных состояний

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

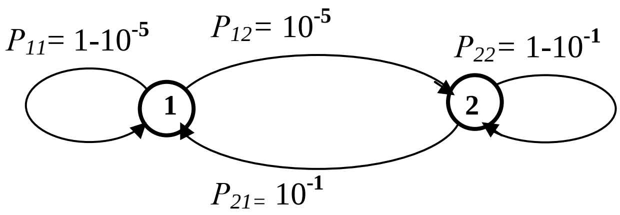


Рис. 1.5. Граф канала

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

Вероятности, составляющие полную группу событий:

$$P_1 + P_2 = 1$$

$$P_{11} + P_{12} = 1 \qquad (1.13)$$

$$P_{22} + P_{21} = 1$$

Вероятности пребывания канала в состояниях 1 и 2:

$$P_1 = P_1 P_{11} + P_2 P_{21};$$

$$P_2 = P_2 P_{22} + P_1 P_{12}. \qquad (1.14)$$

Выразим в (1.28) P_2 через P_1:

$(1\text{-}P_1)P_{21} + (1\text{-}P_{12})P_1 = P_1$, откуда

$$P_1 = P_{21} / (P_{12} + P_{21}). \qquad (1.15)$$

Аналогично, для состояния 2 получим

$$P_2 = P_{12} / (P_{12} + P_{21}). \qquad (1.16)$$

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

$$P_e = P_2 P^{(2)}_e = P^{(2)}_e P_{12} / (P_{12} + P_{21}) \qquad (1.17)$$

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

$$\bar{l}_2 = \sum_{i=1}^{\infty} iP_2(i) \text{, где} \qquad (1.18)$$

$P_2(i) = P_{22}^{i-1} P_{21}$ - вероятность того, что возникшее плохое состояние канала будет длиться в течение i символов. Тогда (1.18) перепишется в виде:

$$\bar{l}_2 = \sum_{i=1}^{\infty} iP_{22}^{i-1} P_{21} = 1 \cdot P_{21} + 2P_{22}P_{21} + 3P_{22}^2 P_{21} + \dots, \qquad (1.19)$$

который представляет собой арифметико-геометрическую прогрессию.

В справочнике [6] приведена сумма такой прогрессии:

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

$$\bar{l}_2 = \frac{1}{P_{22}}\frac{P_{21}P_{22}}{(1-P_{22})^2} = \frac{1}{P_{21}} \qquad (1.20)$$

Аналогично найдем среднюю длину интервала между ошибками:

$$\bar{l}_1 = \sum_{i=1}^{\infty} iP_{11}^{(i-1)}P_{12} = \frac{1}{P_{12}} \qquad (1.21)$$

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

1.3.3.2 Характерные искажения в канале

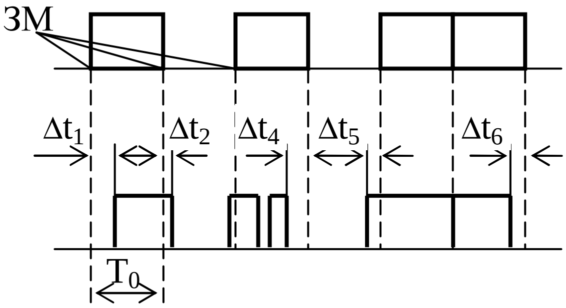


Рис. 1.6. Искажения сигналов

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.

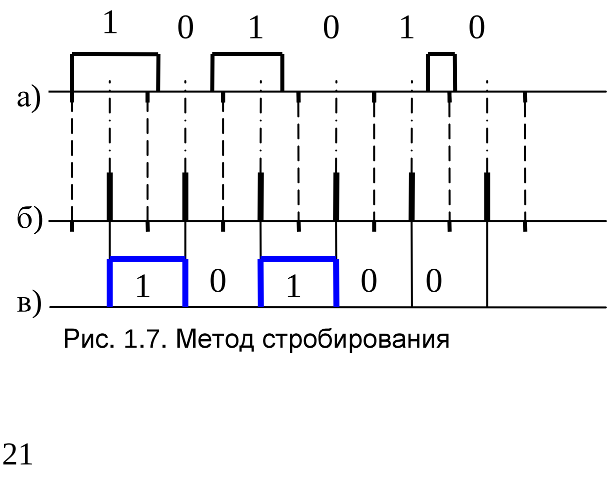


Рис. 1.7. Метод стробирования

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

1) *Метод стробирования*

Верхняя диаграмма рис.1.7 *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, возможна как в результате дроблений, так и в результате краевых искажений.

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. 1.8).

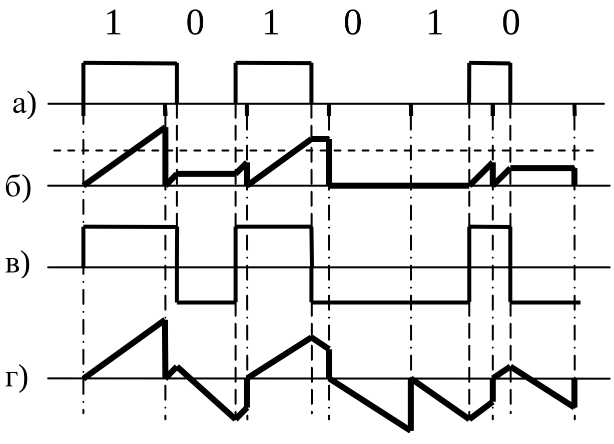


Рис. 1.8 Метод интегрирования

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

1.3.4 Межсимвольная интерференция

На рис. 1.9, *а* представлены фильтрующие элементы типичной системы цифровой связи. В системе — передатчике, приемнике и канале — используется множество разнообразных фильтров (и реактивных элементов, таких как емкость и индуктивность). В передатчике информационные символы, описываемые как импульсы или уровни напряжения, модулируют импульсы, которые затем фильтруются для согласования с определенной полосой частот. В низкочастотных системах канал (кабель) имеет распределенное реактивное сопротивление, искажающее импульсы. Некоторые полосовые системы, такие как беспроводные, являются, по сути, каналами с замираниями, которые проявляют себя как нежелательные фильтры, также искажающие сигнал. Если принимающий фильтр настраивается на компенсацию искажения, вызванного *как* передатчиком, *так и* каналом, он часто называется *выравнивающим* (equalizing filter) или *принимающим/выравнивающим* (receiving/equalizing). На рис. 1.9, *б* приведена удобная модель системы, объединяющая все эффекты фильтрации в одну общесистемную передаточную функцию:

$$H(f) = H_i(f)\, H_c(f)\, H_r(f). \qquad (1.22)$$

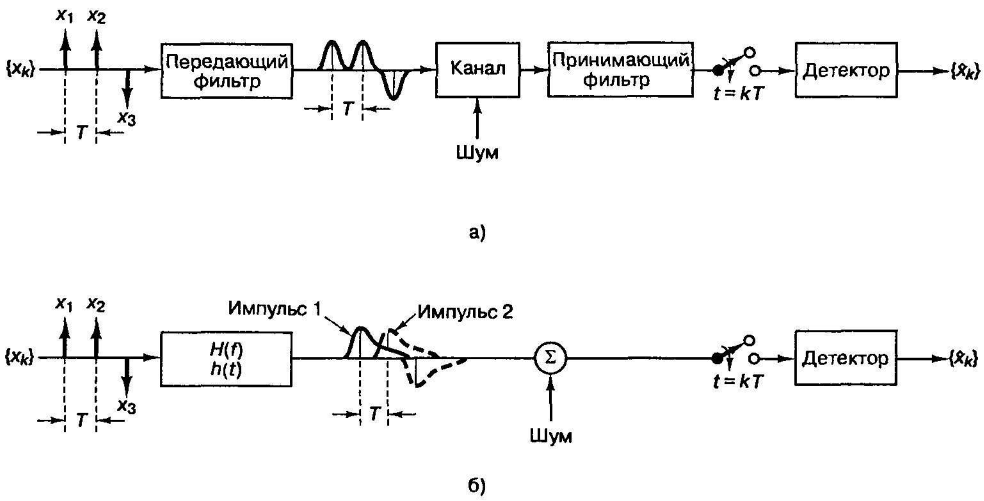


Рис.1.9. Межсимвольная интерференция в процессе детектирования: *а)* типичная низкочастотная цифровая схема; *б)* эквивалентная модель

Здесь $H_i(f)$ характеризует передающий фильтр, $H_c(f)$ – фильтрацию в канале, а $H_r(f)$ - принимающий/выравнивающий фильтр. Таким образом, характеристика $H_i(f)$ представляет передаточную функцию всей системы, отвечающую за все этапы фильтрации в различных местах цепочки передатчик-канал-приемник. В бинарной системе, использующей какую-нибудь распространенную кодировку PCM, например NRZ-L, детектор принимает решение относительно значения символа путем сравнения выборки принятого импульса с порогом. Например, детектор, изображенный на рис. 1.9, решает, что была послана двоичная единица, если принятый импульс положителен, или двоичный нуль — в противном случае. Вследствие системной фильтрации принятые импульсы могут перекрываться, как показано на рис. 1.9, *б*. Хвост импульса может "размазываться" на соседний интервал передачи символа, мешая, таким образом, процессу детектирования и повышая вероятность появления ошибки; подобный процесс получил название *межсимвольной интерференции* (МСИ) или intersymbol interference — ISI. Даже при отсутствии шумов воздействие фильтрации и искажение, вызванное каналом, приводят к возникновению МСИ. Иногда функция $H_c(f)$ задается, и задача состоит в определении $H_i(f)$ и $H_r(f)$ минимизирующих МСИ на выходе выравнивающего фильтра.

Исследованием проблемы задания формы принятого импульса с тем, чтобы предотвратить появление МСИ на детекторе, долгое время занимался Найквист [5]. Он показал, что минимальная теоретическая ширина полосы системы, требуемая для детектирования R_s символов/секунду без МСИ, равна $R/2$ Гц. Это возможно, если передаточная функция системы $H(f)$ имеет прямоугольную форму, как показано на рис. 1.10, *а*. Для низкочастотных систем с такой $H(f)$, что односторонняя ширина полосы фильтра равна $1/2T$ *(идеальный фильтр Найквиста)*, импульсная характеристика функции $H(f)$, вычисляемая с помощью обратного преобразования Фурье, имеет вид $h(t)$ = sinc (t/T); она показана на рис. 1.10, *б*. Импульс, описываемый функцией sinc (t/T), называется *идеальным импульсом Найквиста;* он имеет бесконечную длительность и состоит из многочисленных лепестков: главного и боковых, именуемых *хвостами.* Найквист установил, что если каждый импульс принятой последовательности имеет вид sinc (t/T), импульсы могут детектироваться без межсимвольной интерференции. На рис. 1.10, *б* показано, как удается обойти МСИ. Итак, имеем два последовательных импульса, $h(t)$ и $h(t - T)$. Несмотря на то что хвосты функции $h(t)$ имеют бесконечную длительность, из рисунка видно, что в момент $t = T$ взятия выборки функции $h(t - T)$ хвост функции $h(t)$ проходит через точку нулевой амплитуды, и подобным образом он будет иметь нулевую амплитуду в моменты взятия выборок всех остальных импульсов последовательности $h(t - кT)$, $к = +1, +2, \ldots$. Следовательно, предполагая идеальную синхронизацию процесса взятия выборок, получаем, что межсимвольная интерференция не будет влиять на процесс детектирования. Чтобы низкочастотная система могла детектировать $1/T$ таких импульсов (символов) в секунду, ширина ее полосы должна быть равна $1/2T$; другими словами, система с шириной полосы $W = 1/2T = R_s/2$ Гц может поддерживать максимальную скорость передачи $2W = 1/T = R_S$ символов/с *(ограничение полосы по Найквисту)* без МСИ. Следовательно, при идеальной фильтрации Найквиста (и нулевой межсимвольной интерференции) максимальная возможная скорость передачи символов на герц полосы, называемая *уплотнением скорости* *передачи символов,* равна 2 символа/с/Гц. Вследствие прямоугольной формы передаточной функции идеального фильтра Найквиста и бесконечной длины соответствующего импульса, подобные идеальные фильтры нереализуемы; реализовать их можно только приближенно.

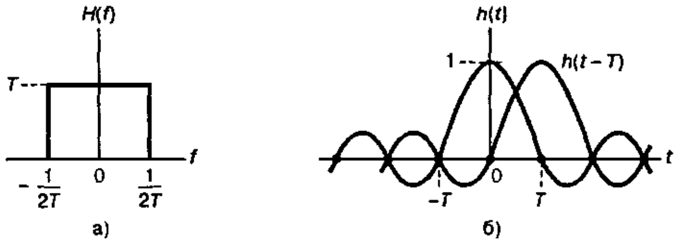


Рис. 1.23. Каналы Найквиста для нулевой межсимвольной интерференции: а) прямоугольная передаточная функция системы H(f); б) принятый импульс h(t) = sinc (t/T)

Стоит отметить, что названия "фильтр Найквиста" и "импульс Найквиста" часто используются для описания обширного класса фильтраций и импульсных форм, удовлетворяющих условию нулевой межсимвольной интерференции в точках взятия выборок. В русскоязычной литературе теорема Найквиста, от которой происходят термины "фильтр Найквиста" и "импульс Найквиста", часто называют теоремой отсчетов или теоремой Котельникова. Фильтр Найквиста — это фильтр, передаточная функция которого может быть представлена прямоугольной функцией, свернутой с любой четно-симметричной частотной функцией. Импульс Найквиста — это импульс, форма которого может быть описана функцией sinc (t/T), умноженной на другую временную функцию. Следовательно, существует бесконечное множество фильтров Найквиста и соответствующих импульсов. В классе фильтров Найквиста наиболее популярными являются фильтры с характеристикой типа приподнятого косинуса или корня из приподнятого косинуса.

Основным параметром систем связи является удельные затраты полосы частот или *эффективность использования полосы*, R/W, измеряемая в бит/с/Гц. Как можно понять из единиц измерения, R/W представляет меру скорости переноса данных на единицу ширины полосы, а значит, показывает, насколько эффективно любой метод передачи сигналов использует ресурс полосы. Поскольку ограничение ширины полосы по Найквисту устанавливает теоретическое максимальное уплотнение скорости передачи символов без межсимвольной интерференции, равное 2 символа/с/Гц, может возникнуть вопрос, можно ли что-то сказать об ограничении величин, измеряемых в бит/с/Гц. О последних ничего нельзя сказать прямо; ограничение связано только с импульсами или символами и возможностью детектирования их амплитудных значений без искажения со стороны других импульсов. При нахождении R/W для любой схемы передачи сигналов необходимо знать, сколько битов представляет каждый символ, что само по себе является темой отдельного рассмотрения. Допустим, сигналы кодируются с использованием *M*-уровневой кодировки PAM. Каждый символ (включающий *k* бит) представляется одной из *M* импульсных амплитуд. Для $k = 6$ бит на символ размер набора символов составляет $M = 2^k = 64$ амплитуды. Таким образом, при 64-уровневой кодировке PAM теоретическая максимальная эффективность использования полосы, не допускающая межсимвольной интерференции, равна 12 бит/с/Гц.

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: $S=\{s_i(t): i=1, 2, ..., M\}$. 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:

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

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

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

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 $s_i = s(iT_s), i = 0, 1, \ldots,$

взятыми через постоянный интервал T_s. Если вся энергия сигнала сосредоточена в пределах полосы W, а $T_s \le 1/2W$, 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 $\mathbf{s} = (s_0, s_1, \ldots, s_{n-1})$ полностью описывает сигнал. Проделав подобную же операцию с наблюдением $y(t)$, приходим к его n–мерному векторному эквиваленту $\mathbf{y} = (y_0, y_1, \ldots, y_{n-1})$, 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*:

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

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).

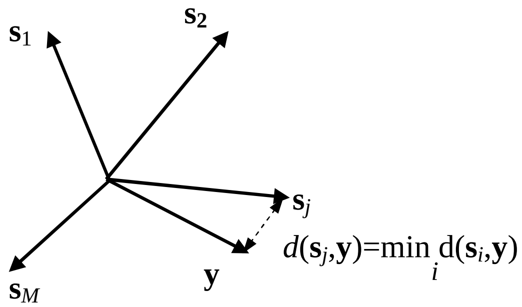


Рис. 2.1. Правило минимума расстояния.

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

$$E = \int_0^T s^2(t)dt \tag{2.4}$$

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

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

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

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

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

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

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*:

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

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

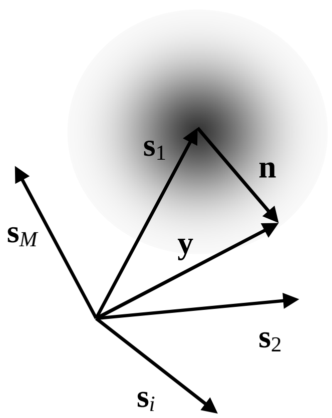


Рис. 2.2. Зашумленное наблюдение и задача выбора сигналов.

более чем явный способ конструирования множества сигналов. Обратимся к рис.2.2, на котором изображены сигнальные вектора. Предположим, что передавался сигнал $\mathbf{s}_1$, и что он подвергся искажению в АБГШ канале, следствием чего служит добавление к $\mathbf{s}_1$ вектора шума $\mathbf{n}$. Гауссовский вектор $\mathbf{n}$ характеризуется симметричным (сферическим) вероятностным распределением, экспоненциально спадающим с увеличением длины вектора $\mathbf{n}$, что очевидным образом следует из (2.1) после удаления из него сигнала (т.е. при подстановки $s(t) = 0$). Следовательно, вектор наблюдения $\mathbf{y} = \mathbf{s}_1 + \mathbf{n}$ будет случайным образом перемещаться вокруг $\mathbf{s}_1$, как это показано на рисунке, и тогда, согласно правилу минимума расстояния (2.3), как только $\mathbf{y}$ окажется ближе к некоторому другому, чем $\mathbf{s}_1$, сигналу, то будет принято ошибочное решение. Для минимизации возникновения ошибки подобного рода следует располагать другие сигналы на максимально большом расстоянии от $\mathbf{s}_1$. Поскольку любой из M сигналов может передаваться равновероятно, т.е. занимать место $\mathbf{s}_1$, то, очевидно, что все расстояния $d(\mathbf{s}_i, \mathbf{s}_j), 1 \le i < j \le M$, 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 сигналов.

2.2. Передача двоичных данных (детерминированные сигналы).

Строгая зависимость качества приема от расстояния между сигналами наиболее наглядно проявляется в простейшей, но и очень типичной связной 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

$$d(\mathbf{s}_0, \mathbf{y}) \underset{\hat{H}_1}{\overset{\hat{H}_0}{\lessgtr}} d(\mathbf{s}_1, \mathbf{y}), \tag{2.9}$$

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

$$z = z_0 - z_1 \underset{\hat{H}_1}{\overset{\hat{H}_0}{\gtrless}} \frac{E_0 - E_1}{2}, \tag{2.10}$$

где корреляция $z_i, i = 0, 1$ каждого из сигналов и наблюдения $y(t)$ определяется соотношением (2.7), а энергия сигнала $E_i = \|\mathbf{s}_i\|^2, i = 0, 1$ устанавливается (2.4). Оптимальные правила различения (2.9)–(2.10) двух сигналов могут быть наглядно объяснены с помощью их геометрической интерпретации. Два сигнальных вектора $\mathbf{s}_0$ и $\mathbf{s}_1$ всегда лежат в сигнальной плоскости *SP*. Вектор наблюдения $\mathbf{y}$ необязательно попадет на эту плоскость, но близость его к одному или другому сигналу определяется близостью к ним проекции $\mathbf{y}'$ вектора $\mathbf{y}$ на *SP* (см. рис. 2.3, *a*). Следовательно, плоскость *SP* может быть поделена на

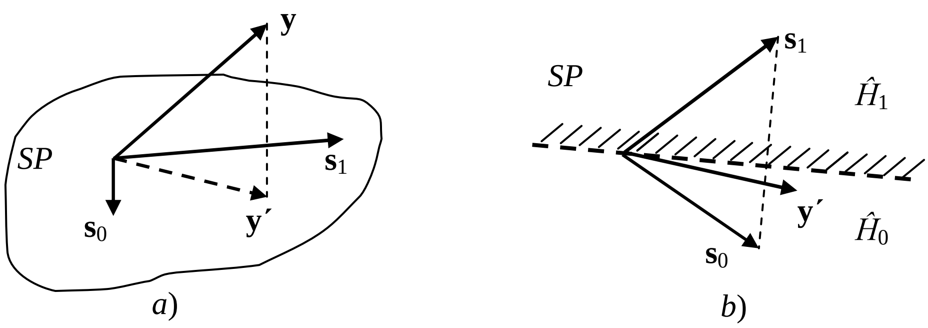


Рис. 2.3. Сигнальная плоскость и полуплоскости решения.

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 $\mathbf{y}'$ на соответствующую полуплоскость (см. рис. 2.3, *b*). Из рис. 2.3, *b* также следует, что вероятность перепутывания сигналов (вероятность ошибки) 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

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

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 процесс 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

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

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 вверху 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

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

где

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

– *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

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

где квадрат интеграла заменен двойным интегралом с раздельными 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

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

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

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

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

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

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 прием P_e не зависит априори от вероятности w поступления в канал сигнала $s_0(t)$ и устанавливается соотношением

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

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

$$z = z_0 - z_1 \underset{\hat{H}_1}{\overset{\hat{H}_0}{\gtrless}} 0 .$$

Очевидно, что для максимизации расстояния между двумя векторами фиксированной длины следует выбирать их *противоположными*, как это показано на рис. 2.4, *a*. 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

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

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 непо-

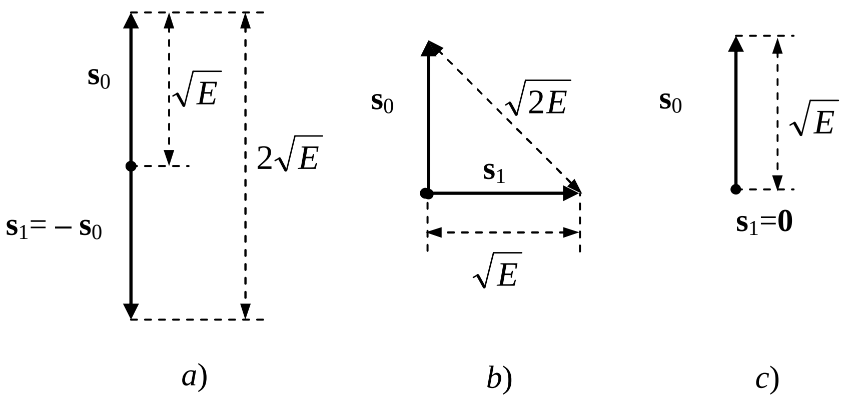


Рис. 2.4. Варианты выбора пар двоичных сигналов.

средственного вычисления корреляции z, а параметр $q = \sqrt{2E/N_0}$ есть ничто иное, как *отношение сигнал-шум* на выходе упомянутых устройств.

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 возможных 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 удается 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. $\cos\alpha = \rho_{01} = 0$, $d(\mathbf{s}_0, \mathbf{s}_1) = \sqrt{2E}$ (см. рис. 2.4, *b*). Подстановка этих значений в (2.18) дает

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

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*) $d(\mathbf{s}_0, \mathbf{s}_1) = \sqrt{E}$, а вероятность ошибки (2.18) становится равной

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

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

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

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

Замечание о записи амплитуды сигнала.

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

$A_{ско}^2$ представляет собой среднюю мощность P, рассеиваемую на сопротивлении 1 Ом, следовательно

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

Заменяя P на E/T, получаем:

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

Конец замечания.

Очевидно, различение таких сигналов должно производится 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.

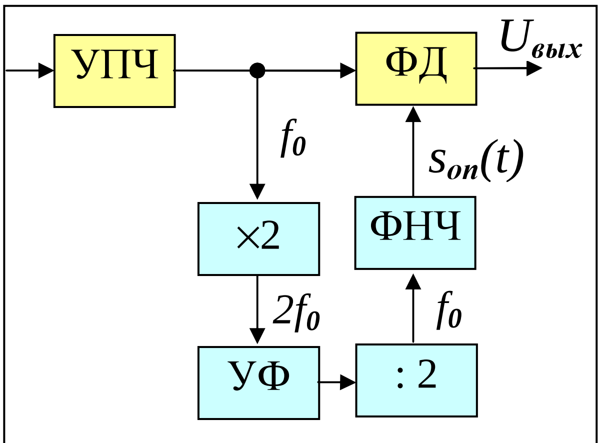


Рис. 2.5. Схема Пистолькорса

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 (×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 (×2). В этой схеме отсутствует делитель 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.

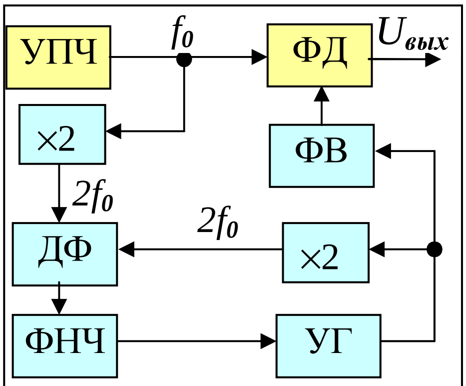


Рис. 2.6. Схема Сифорова

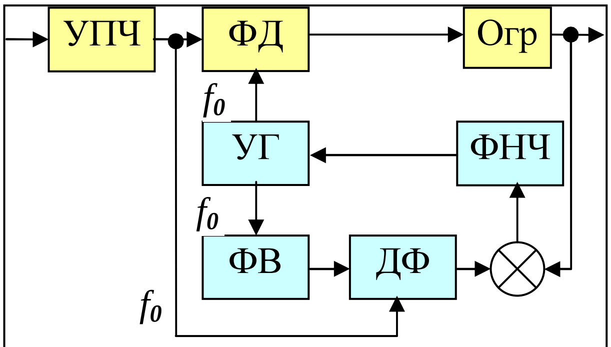


Рис. 2.7. Схема Костаса

Предложенная в 1957 году американским ученым Д. Костасом схема формирования опорного напряжения показана на рис. 2.7. Все элементы схемы работают на частоте f_0, а фазовая манипуляция снимается умножением продетектированного 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: текущий передаваемый символ x_k меняет свою фазу относительно предыдущего при информационном символе m_k “1” (M-код) или “0” (S-код).

$$x_k = x_{k-1} \oplus m_k \quad \text{или}$$
$$x_k = \overline{x_{k-1} \oplus m_k} \qquad (2.23)$$

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

В общем случае M-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.

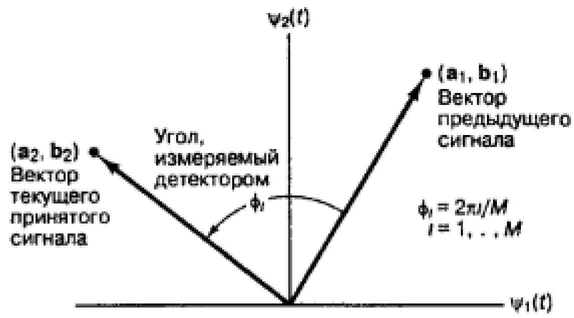


Рис. 2.8. Сигнальное пространство для M-ичной ОФМ

Если определить набор сигналов ОБФМ следующим образом

$$x_1(t) = A\cos(\omega_0 t + \varphi) \qquad 0 \le t \le T,$$
$$x_2(t) = A\cos(\omega_0 t + \varphi \pm \pi) \qquad 0 \le t \le T,$$

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

$$s_1(t) = (x_1, x_1) \quad \text{или} \quad (x_2, x_2) \quad 0 \le t \le 2T,$$
$$s_2(t) = (x_1, x_2) \quad \text{или} \quad (x_2, x_1) \quad 0 \le t \le 2T \qquad (2.24)$$

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.

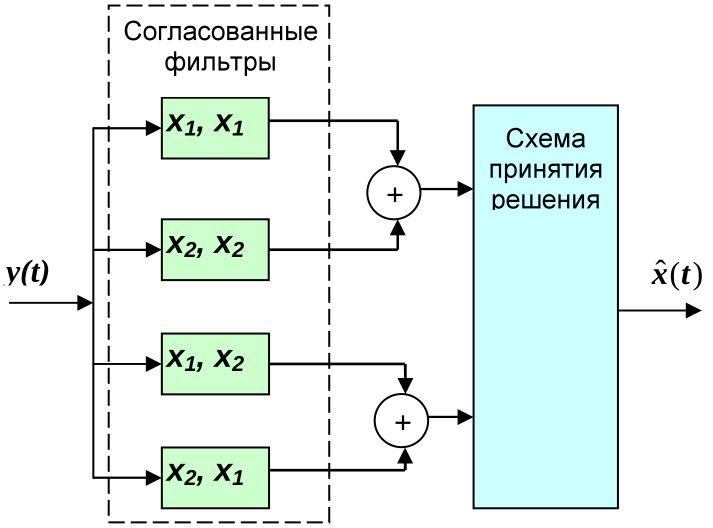


Рис. 2.9. Дифференциально-когерентный приемник ОБФМ

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

$$P_E = 0{,}5\exp(-E_b/N_0) \qquad (2.25)$$

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

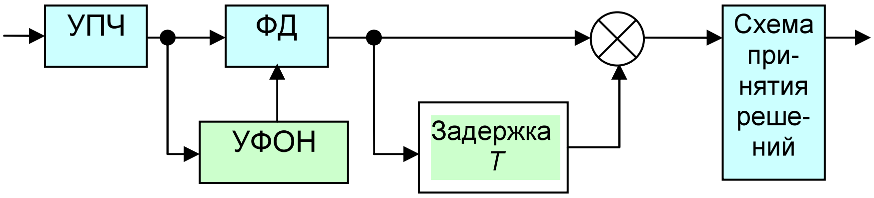


Рис. 2.10. Квазикогерентный приемник ОБФМ.

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

$$P_E = 2P_{1E}\,(1 - P_{1E}) + P_{ОП},$$

где $P_{1E} = Q(h\sqrt{2})$ – вероятность ошибки приема одного двоичного сигнала, $P_{ОП}$ – 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

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

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

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

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

$$P_E \approx \frac{1}{h\sqrt{\pi}}\exp\left(-h^2\right). \qquad (2.27)$$

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.

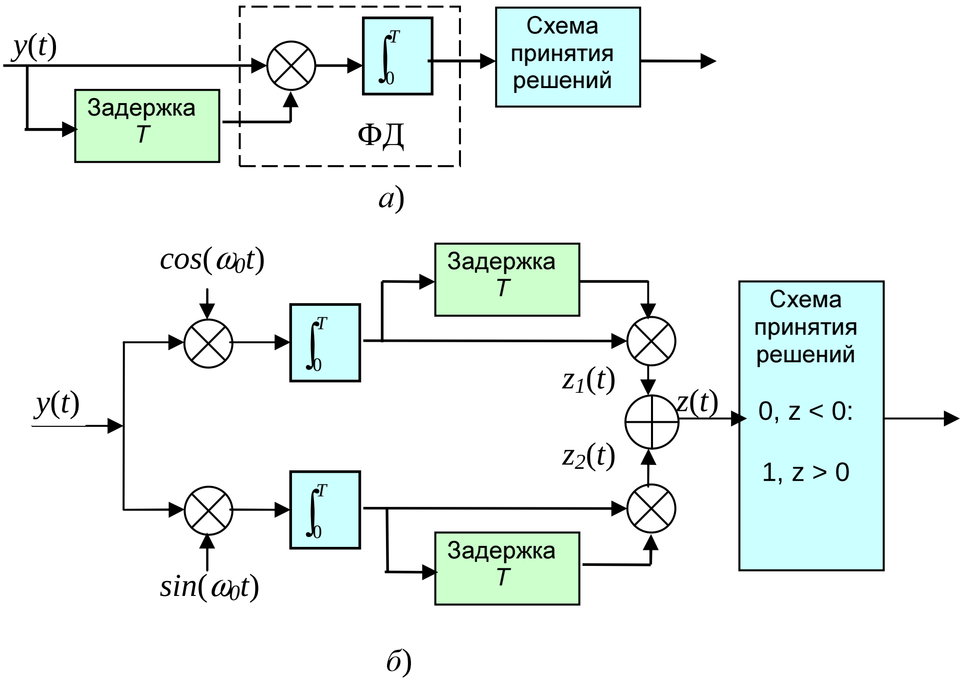


Рис. 2.11. Приемники ОФМ-сигналов: а) автокорреляционный; б) квадратурный

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

2.3. Передача М–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 $s_j(t),\ j = 2, 3, \ldots, M$, 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 $z_1 \geq z_i - \frac{E_i - E_1}{2},\ i = 1, 2, \ldots, M$. 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 $s_j(t),\ j = 2, 3, \ldots, M$, 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.

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

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 (2.17) 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

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

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

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

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

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

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 $\overline{E} = \sum_{i=1}^{M} E_i / M = const$, 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 $E = E_i = const$, $i = 1, 2, \ldots, M$ 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

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

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

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

где $S_I(t) = S(t)\cos\gamma(t)$ и $S_Q(t) = S(t)\sin\gamma(t)$ – квадратурные компоненты сигнала. Поскольку и $S(t)$, 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, $n_s = 2W_t T_t$.

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 $\overline{E} = \min$ (объемная упаковка), либо равной для всех сигналов $E = \min$ (сферическая упаковка).

Простейшая версия этой задачи ($n_S = 1$) отвечает случаю амплитудной (АМ) модуляции (простейший вариант которой с $M = 2$ – БАМ). Альтернативным наименованием АМ служит *импульсно–кодовая модуляция* (ИКМ)).

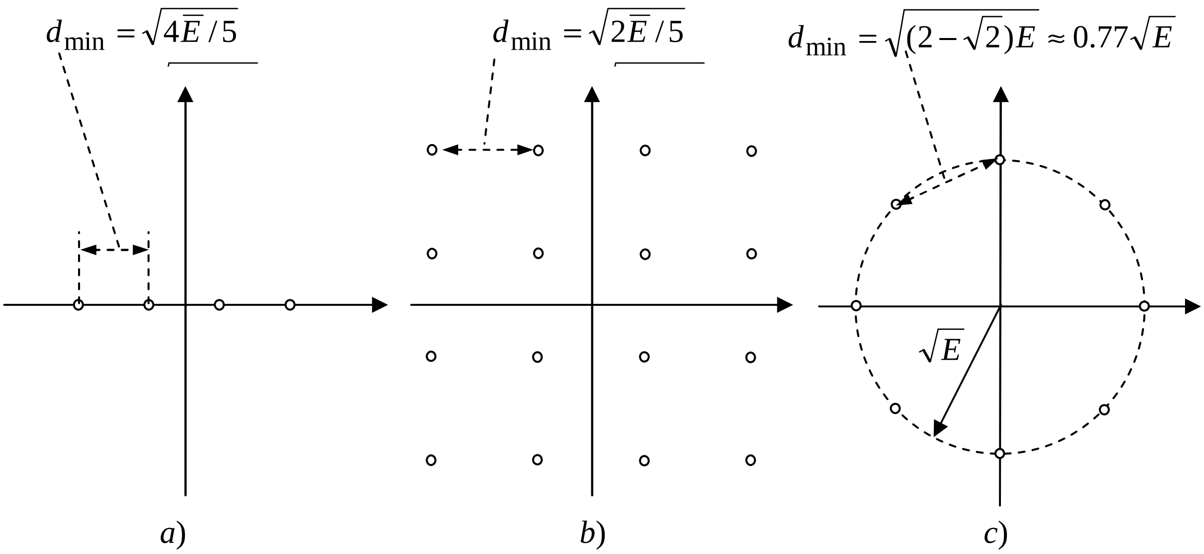


Рис. 2.12. Одно- и двумерные созвездия: 4-АМ (*a*), 16-КАМ (*b*), и 8-ФМ (*c*).

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, *a*).

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 теоретически, 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 $(E_i = E, i = 1, 2, \ldots, M)$, 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

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

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 раздельным 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

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

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

$$d_{\min}^2 \le \frac{2M}{M-1} E\,. \tag{2.33}$$

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 $\mathbf{u}_i, i = 1, 2, \ldots, M$, имеющих попарно нулевое скалярное произведение и единичную длину: $(\mathbf{u}_i, \mathbf{u}_j) = \delta_{ij}, i, j = 1, 2, \ldots, M$, где $\delta_{ij} = 0, i \ne j; \delta_{ij} = 1, i = j$ – дельта-функция Кронекера. Такие вектора, называемые *ортонормированными*, существуют в любом векторном пространстве, размерность которого не меньше, чем M. Сформируем теперь M новых векторов $\mathbf{v}_i, i = 1, 2, \ldots, M$, 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

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

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:

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

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

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

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

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

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

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

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.

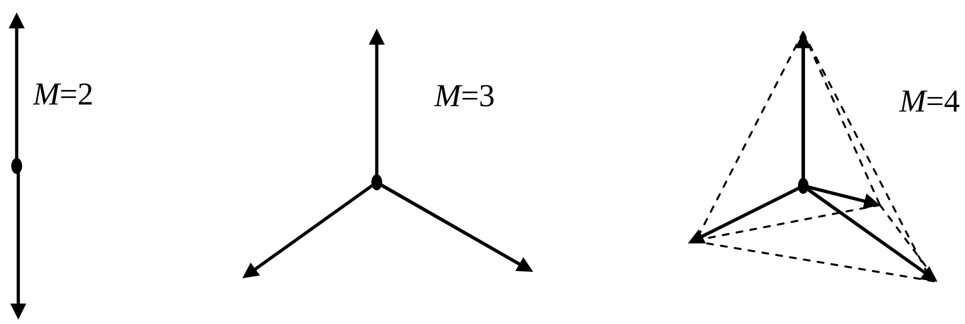


Рис. 2.13. Примеры симплексных сигналов.

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:

$$P_{e\ ort} \le (M-1)\,Q\left(\sqrt{\frac{E}{N_0}}\right),\ \frac{2E}{N_0} >> 1. \qquad (2.38)$$

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 $M >> 1$ 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 $s(t) = S(t)\cos[2\pi f_0 t + \gamma(t)]$ 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, когда 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

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

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

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

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

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:

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

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

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

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

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 $\dot{S}_j(t)$:

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

где $\dot{\rho}_{ij}$, как и в (2.16), коэффициент корреляции, но адаптированный к комплекснозначным сигналам, т.е. комплексным огибающим

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

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 $\operatorname{Re}(\dot{\rho}_{ij}) = \rho_{ij}$, 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

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

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

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

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

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

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

Теперь квадрат расстояния, приведенный выше, будет записан, как

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

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 $|\rho_{ij}|$ 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

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

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 $2W_tT_t$, как это было в случае видеосигналов, до значения W_tT_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 $2W_tT_t$, то в нем можно расположить только W_tT_t ортогональных плоскостей, а, значит, и W_tT_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 $\dot{Y}(t)$: $y(t) = \mathrm{Re}[\dot{Y}(t)\exp(j2\pi f_0 t)]$. 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: $d^2(\dot{\mathbf{S}}_{i\,\varphi}, \dot{\mathbf{Y}}) = \|\dot{\mathbf{Y}}\|^2 + 2E - 2\,\mathrm{Re}[\dot{z}_i(\varphi_i)]$, 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

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

означающее, что j–й сигнал считается принятым, если модуль корреляции между его полностью 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.

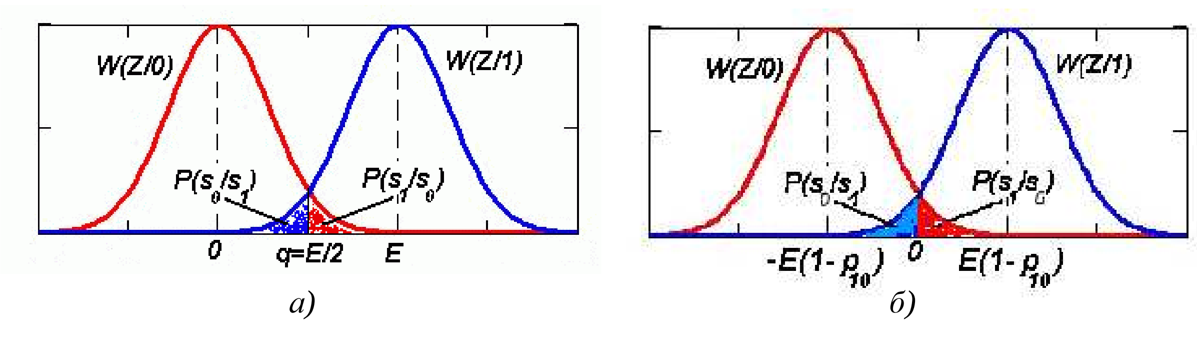


Рис. 2.14. Плотности распределения сигналов на входе решающей схемы при когерентном приеме: а) – БАМ; б) – БЧМ и БФМ

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 БАМ представлена на рис. 2.15, а ее распределения – на рис. 2.16.

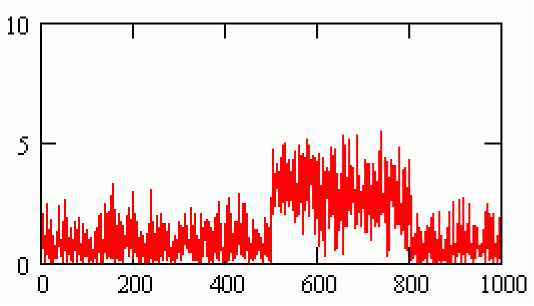


Рис. 2.15. Огибающая сигнала БАМ.

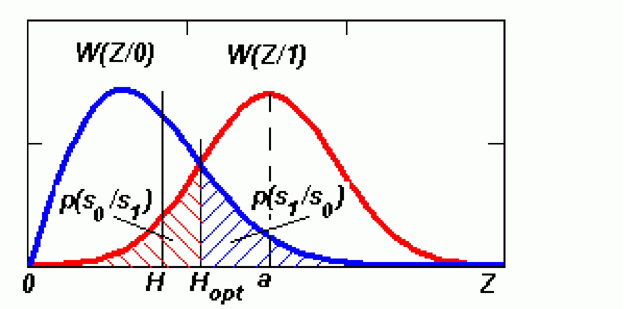


Рис. 2.16. Распределения огибающей БАМ.

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, $I_0(.)$ – 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

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

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

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: $H = U_{пор}/\sigma$.

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 $W(Z/0)|_{Z=Hopt} = W(Z/1)|_{Z=Hopt}$ 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. Заметим, что $H_{opt}=\frac{U_{Порт}}{\sigma}=\frac{U_{Порт}}{S_0}\frac{S_0}{\sigma}=\frac{U_{Порт}}{S_0}\sqrt{2}h.$

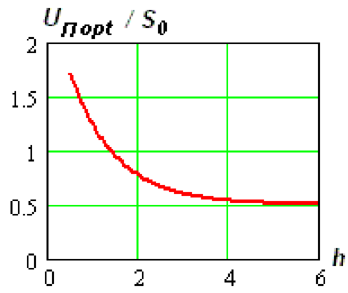


Рис.2.17. Зависимость оптимального порога от отношения сигнал/шум

При достаточно большом ($h \geq 3$) отношении сигнал/шум величина оптимального порога $U_{Порт}/S_0$ 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:

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

Используем эту аппроксимацию для вычисления

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

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

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

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

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

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

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

Полная вероятность ошибки

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

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

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

Частотно-манипулированные сигналы

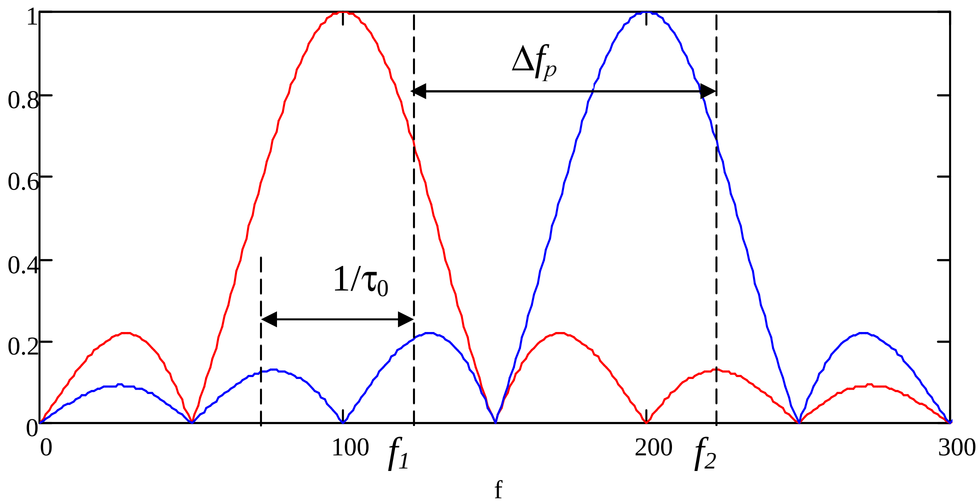


Рис. 2.18. Спектр сигналов БЧМ

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

Оптимальный в смысле экономии частотного ресурса разнос частот Δf_{popt} = 0,75/Т. Ниже этого значения различение сигналов ухудшается, его увеличение не приводит к росту помехоустойчивости.

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

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

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.

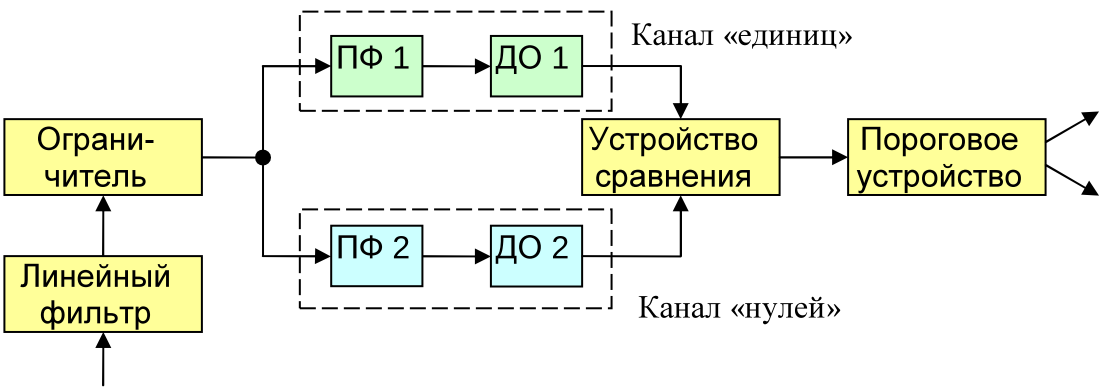


Рис. 2.19. Приемник ЧМ-сигналов с двумя полосовыми фильтрами

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

$$P(U_{Ш} \geq U_{СШ}) = \int_{U_{СШ}}^{\infty} W(U_{Ш}) dU_{Ш} \tag{2.56}$$

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

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

$$P(s_0/s_1) = \int_0^{\infty} \int_{U_{СШ}}^{\infty} W(U_{СШ}) W(U_{Ш}) dU_{Ш} dU_{СШ}, \tag{2.57}$$

где

$W(U_{СШ})$ – распределение Релея-Райса,

$W(U_{Ш})$ - распределение Релея. Используя $v=U_{СШ}/\sigma$, $z=U_{Ш}/\sigma$, $a = S_0/\sigma$, получим:

$$\begin{gathered} P(s_0/s_1) = \int_0^{\infty} \int_{U_{СШ}}^{\infty} \frac{U_{СШ}}{\sigma^2} \exp\left(-\frac{U_{СШ}^2 + S_0^2}{2\sigma^2}\right) I_0\left(\frac{U_{СШ} S_0}{\sigma^2}\right) \frac{U_{Ш}}{\sigma^2} \exp\left(-\frac{U_{Ш}^2}{2\sigma^2}\right) dU_{Ш} dU_{СШ} = \\ \int_0^{\infty} \int_{U_{СШ}/\sigma}^{\infty} v \exp\left(-\frac{v^2 + a^2}{2}\right) I_0(av) z \exp\left(-\frac{z^2}{2}\right) dv dz; \end{gathered} \tag{2.58}$$

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

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

2.4.2. Сравнение когерентного и некогерентного способов приема бинарных сигналов

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

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

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

Для отличия обозначим его как $h_{0нкг}$.

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

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

который быстро сходится при x > 1:

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

А. Амплитудная манипуляция

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

$$\begin{cases} P_{ЕКГ} = \left\{Q(h_0/\sqrt{2})\right\} \approx \dfrac{1}{h_0\sqrt{\pi}} e^{-h_0^2/4} \\ P_{ЕНКГ} \approx 0{,}5 e^{-h_{0нкг}^2/4} \end{cases}$$

Равные энергетические затраты

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

$$\frac{P_{\text{Е НКГ}}}{P_{\text{Е КГ}}} = \frac{\sqrt{\pi}h_0}{2}.$$

Проигрыш НКГ приема сигналов БАМ когерентному пропорционален h_0.

Равные ошибки

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

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

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

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 $h_0 > 3$:

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

Равные энергетические затраты

$h_{0нкг} = h_0$. Из (2.58) и (2.59) следует

$$\frac{P_{\text{Е нкг}}}{P_{\text{Е опт}}} = \sqrt{\frac{\pi}{2}}h_0 \approx 1,26h_0$$

Равные ошибки

$$\frac{h_{0\,нкг}^2}{h_0^2} = \frac{1+2\ln\left(1,26h_0\right)}{h_0^2}$$

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_tT_t системы. Первоначально 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, *а*, которые совместно занимают весь временной интервал T_t. При длительности сигнала не более, чем $T = T_t / M$, и временном сдвиге между соседними сигналами не менее, чем длитель-

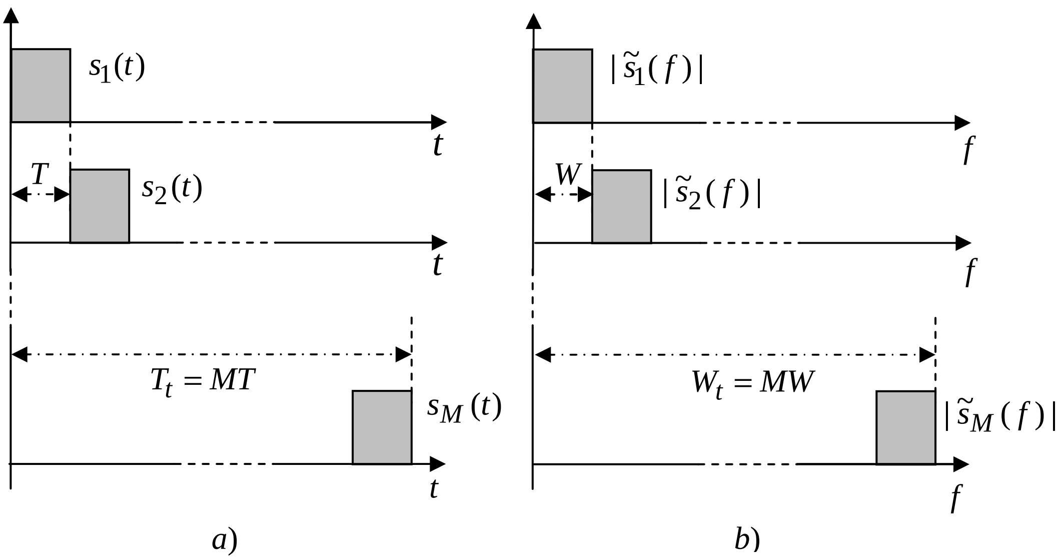


Рис. 2.20. Ортогональные по времени (*a*) и частоте (*b*) сигналы.

ность 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 $M >> 1$ 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.

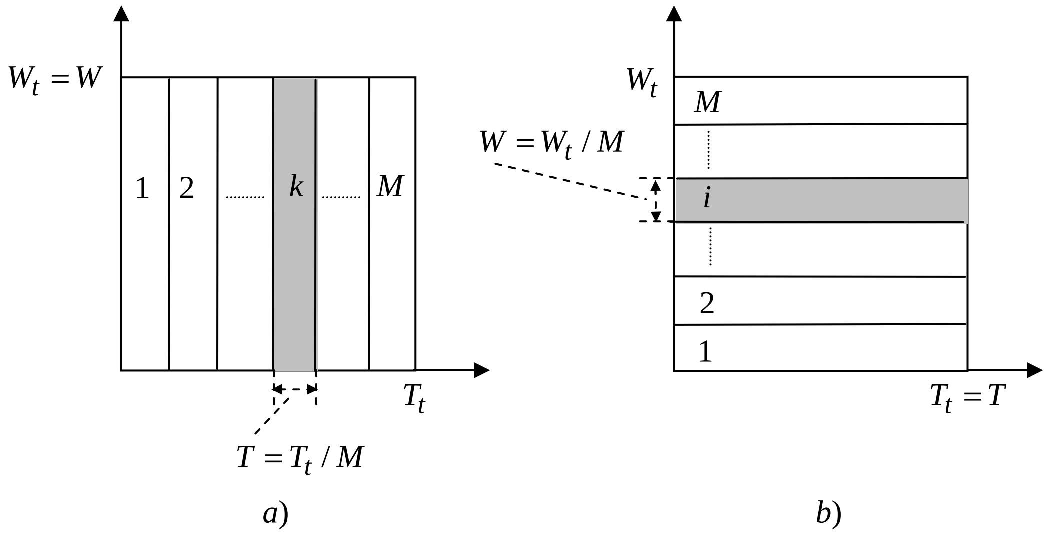


Рис. 2.21. Распределение ресурса при временном (a) и частотном (b) ортогональном кодировании.

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 $\nu = M >> 1$. 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:

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

что позволяет перевести ранее представленный метод в частотную область (см. Рис. 2.20, *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. 2.21, *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 $\nu = 1$ 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 аппаратурной 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. Образуем 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 соответствие с *кодовой последовательностью* (или просто *кодом*) бинарных символов $a_{k,i} = \pm 1$, где $k = 1, 2, \ldots, M$, а второй подстрочный символ отвечает номеру чипа (дискретному времени) $i = 0, 1, \ldots, N-1$. Тогда версия подобного сигнала на видеочастоте может быть записана в виде

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

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

$$(\mathbf{s}_k, \mathbf{s}_l) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} a_{k,\,i} a_{l,\,j} \int_0^T s_0(t - i\Delta) s_0(t - j\Delta) dt\,. \tag{2.63}$$

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

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

Последнее соотношение связывает скалярное произведение сигналов (2.62) со скалярным произведением N–мерных векторов, отвечающих кодовым последовательностям $\mathbf{a}_k = (a_{k,\,0}, a_{k,\,1}, \ldots, a_{k,\,N-1})$. 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 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. $M \equiv 0 \bmod 4$, где используемое сравнение $a \equiv b \bmod c$ 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:

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

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 $M = 2^m$ $(2, 4, 8, 16, 32, \ldots)$. 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

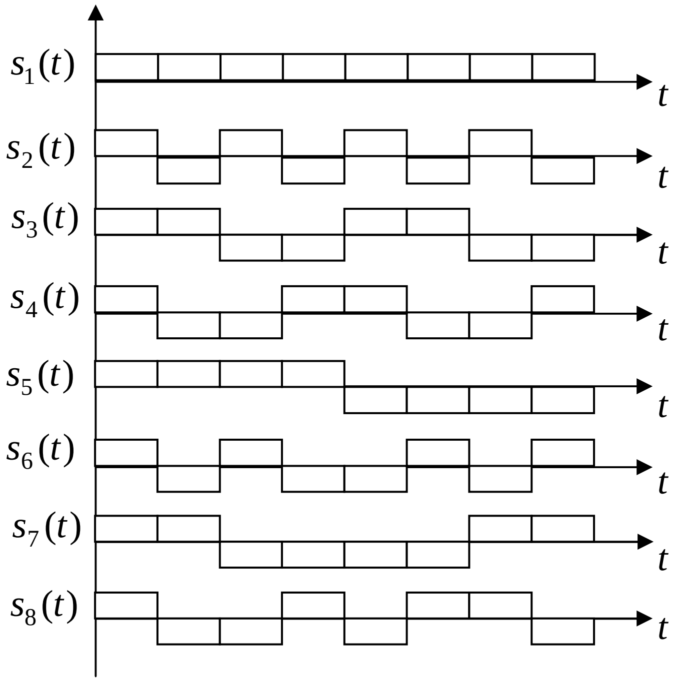


Рис. 2.21. Функции Уолша на основе H_8.

$\mathbf{H}_8$.

Рис. 2.21 иллюстрирует тот факт, что в рамках данного метода построения ортогональных 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).

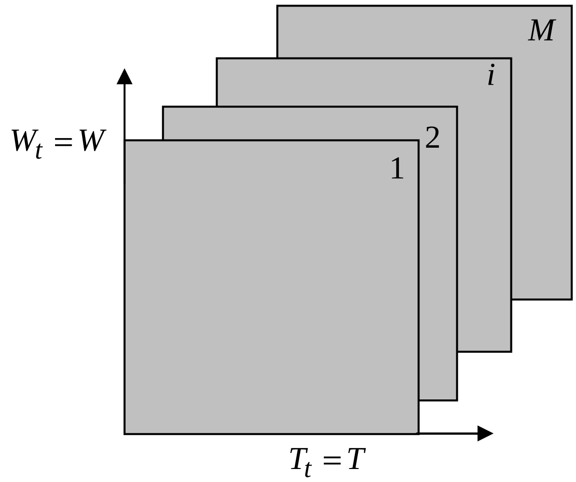


Рис. 2.21. Распределение ресурса при ортогонализации.

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 $W_t T_t >> 1$, 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.

Замечание относительно определения отношения сигнал/шум.

Известно, что в аналоговой связи критерием качества является отношение средней мощности сигнала к средней мощности шума (S/N или SNR). Аналоговый сигнал можно представить как сигнал, имеющий бесконечную длительность, который не требуется разграничивать во времени. Неограниченно длительный аналоговый сигнал содержит бесконечную энергию; следовательно, использование энергии — это не самый удобный способ описания характеристик такого сигнала. Значительно более удобным параметром для аналоговых волн является мощность (или скорость доставки энергии).

В то же время в системах цифровой связи мы передаем (и принимаем) символы путем передачи некоторого сигнала в течение конечного промежутка времени, времени передачи символа *Ts*. Сконцентрировав внимание на одном символе, видим, что мощность (усредненная по времени) стремится к нулю. Значит, для описания характеристик цифрового сигнала мощность не подходит. Для подобного сигнала нам нужна метрика, "достаточно хорошая" в пределах конечного промежутка времени. Другими словами, энергия символа (мощность, проинтегрированная по *Ts)* — это гораздо более удобный параметр описания цифровых сигналов.

То, что цифровой сигнал лучше всего характеризует полученная им энергия, еще не дает ответа на вопрос, почему E_b/N_0 — это естественная метрика для цифровых систем, так что продолжим. Цифровой сигнал — это транспортное средство, представляющее цифровое сообщение. Сообщение может содержать один бит (двоичное сообщение), два (четверичное), 10 бит (1024-ричное). В аналоговых системах нет ничего подобного такой дискретной структуре сообщения. Аналоговый информационный источник — это бесконечно квантованная непрерывная волна. Для цифровых систем критерий качества должен позволять сравнивать одну систему с другой на битовом уровне. Следовательно, описывать цифровые сигналы в терминах *S/N* практически бесполезно, поскольку сигнал может иметь однобитовое, 2-битовое или 10-битовое значение. Предположим, что для данной вероятности возникновения ошибки в цифровом двоичном сигнале требуемое отношение *SIN* равно 20. Будем считать, что понятия сигнала и его значения взаимозаменяемы. Поскольку двоичный сигнал имеет однобитовое значение, требуемое отношение *S/N* на бит равно 20 единицам. Предположим, что наш сигнал является 1024-ричным, с теми же 20 единицами требуемого отношения *S/N*. Теперь, поскольку сигнал имеет 10-битовое значение, требуемое отношение *S/N* на один бит равно всего 2. Возникает вопрос: почему мы должны выполнять такую цепочку вычислений, чтобы найти метрику, представляющую критерий качества? Почему бы сразу не выразить метрику через то, что нам действительно надо, — параметр, связанный с энергией на битовом уровне, E_b/N_0.

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 однобитовыми противоположными сигналами, т.е. согласно рис. 2.4, *a* $d^2_{\min,\, u} = 4E_b$, где второй индекс обозначает отсутствие кодирования (uncoded).

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*: $d^2_{\min,\, ort} = 2mE_b$. Очевидно, что выигрыш G_a ортогональных сигналов в величине минимума квадрата расстояния у варианта отсутствия кодирования составляет $m/2$.

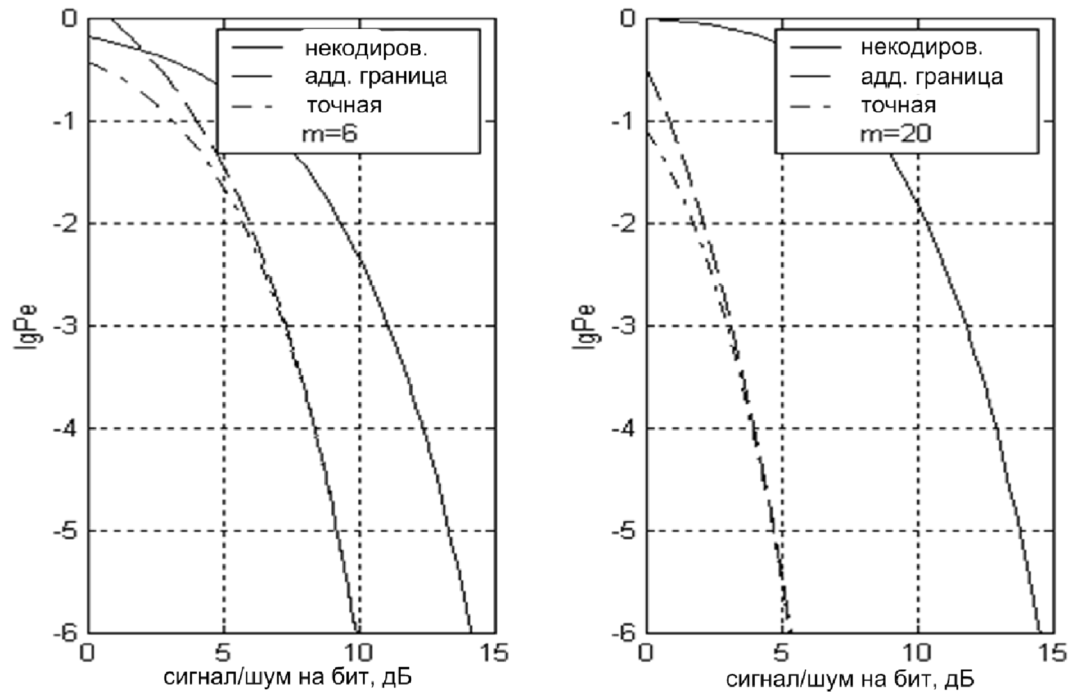


Рис. 2.21. Сравнение вероятности ошибки для некодированной передачи и ортогональной сигнализации

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]:

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

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 $P_e \le 10^{-3}$). 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 в 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 $m = RT_t$ бит за выделенный интервал времени T_t. Ясно, что ортогональное кодирование битовых блоков такой длины может быть осуществлено с помощью $M = 2^m = 2^{RT_t}$ сигналов, обеспечивая асимптотический выигрыш от кодирования $G_a = m/2 = RT_t / 2$. Тогда спектральная эффективность R/W_t системы, т.е. скорость, приходящаяся на 1 Гц полосы, использующей ортогональные сигналы, будет

$$\frac{R}{W_t}=\frac{RT_t}{2^{RT_t}}=\frac{2G_a}{2^{2G_a}}, \tag{2.66}$$

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

Обратимся к числовому примеру.

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 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 $\Delta f_Д$ 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.

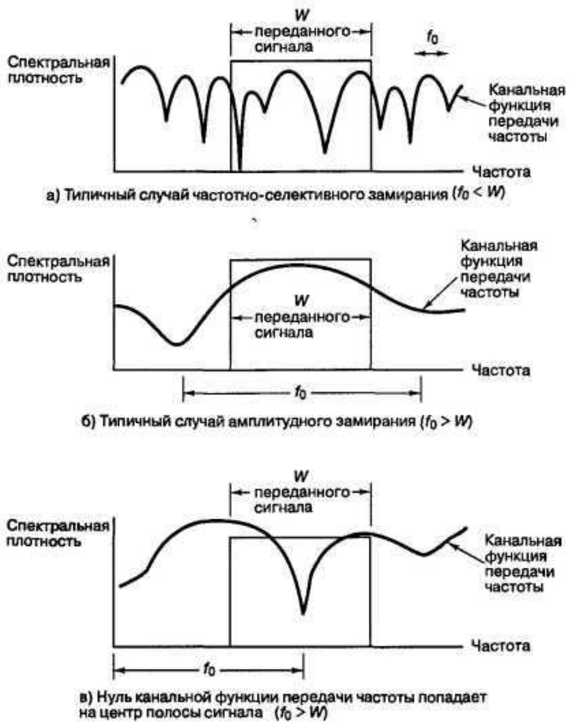


Рис. 3.1. Коэффициент передачи канала и спектр сигнала

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, б видно, что воздействие на все спектральные компоненты будет приблизительно равным. Амплитудное замирание не привносит 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, видеоимпульс 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

1) $\rho(\tau) = \exp\left(-\frac{|\tau|}{\tau_{фл}}\right)$,

2) $\rho(\tau) = \exp\left(-\frac{\tau^2}{2\tau_{фл}^2}\right)$,

где $\tau_{фл}$ – 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) $\tau_k = \tau_{фл}$ и $\rho(\tau_k) \approx 0{,}4$;

2) $\tau_k = \tau_{фл}\sqrt{\pi}/2 \approx 1{,}25\tau_{фл}$ и $\rho(\tau_k) \approx 0{,}45$.

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

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

$$s_x(t) = \sum_k s_{ik}(t) = \sum_k S_{ik}(t)\, f_i\left[t - \tau_{зk}(t)\right] \tag{3.1}$$

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

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

$\tau_{зk}$ – запаздывание посылки k-го луча;

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

$$f_i(t) = \begin{cases} \sin(\omega_0 t + \varphi_i), & 0 \le t \le T \\ 0, & \text{для других } t \end{cases}$$

Принятый сигнал (3.1) имеет две компоненты:

$$s_x(t) = s_p(t) + s_{фл}(t), \tag{3.2}$$

$s_p(t)$ – регулярная, $s_{фл}(t)$ – случайная.

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

Релея-Райса

$$W(S_X) = \frac{S_X}{\sigma_{фл}^2}\exp\left(-\frac{S_P^2 + S_X^2}{2\sigma_{фл}^2}\right) I_0\left(\frac{S_P S_X}{\sigma_{фл}^2}\right), \quad \left\{\overline{S_X^2} = 2\sigma_{фл}^2\right\} \tag{3.3}$$

и Накагами (m-распределение) – более общий закон

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

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

$$s_x(t) = \alpha S_0 \sin(\omega_0 t + \varphi_x), \quad 0 \le t \le T \tag{3.5}$$

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

Тогда

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

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 $\mathrm{M}[P_E(h_x)]$ ищем обычным образом:

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

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

$$h_x^2 = \frac{S_x^2 TW}{2N_0 W} = \frac{S_x^2 B_c}{2\sigma_ш^2} = \frac{P_{cx}}{P_ш} B_c \text{, или}$$

$$h_x = \frac{S_x \sqrt{B_c}}{\sigma_ш \sqrt{2}} = aS_x, \qquad (3.8)$$

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

Распределение случайной амплитуды запишем в виде:

$$W(S_x) = \frac{S_x}{\sigma_{фл}^2} \exp\left(-\frac{S_x^2}{2\sigma_{фл}^2} \right), \qquad (3.9)$$

где $\overline{P_{фл}} = 0,5\overline{S_{фл}^2} = 2\sigma_{фл}^2$ - 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: $h_x = aS_x$,

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

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

$$W(h_x) = \frac{h_x}{a\sigma_{фл}^2} \exp\left(-\frac{h_x^2}{2a^2\sigma_{фл}^2} \right) \frac{1}{a} = \frac{2h_x}{\overline{\overline{h_x^2}}} \exp\left(-\frac{h_x^2}{\overline{\overline{h_x^2}}} \right)$$
$$\overline{h_x^2} = 2a^2\sigma_{фл}^2 = 2\frac{B_c \sigma_{фл}^2}{\sigma_ш^2 \cdot 2} = B_c \frac{\sigma_{фл}^2}{\sigma_ш^2} = h^2 \qquad (3.10)$$

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

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

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

Когерентный прием двоичных сигналов

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

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

Интеграл (3.12) берем по частям, воспользовавшись правилом $\int udv = uv - \int vdu$.

В нашем случае

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

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

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

При вычислении (3.13) использован табличный интеграл [6]

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

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

$$\overline{P_E} \approx 1/2\gamma_c^2 h^2 \tag{3.14}$$

и для различных типов манипуляции получаем:

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

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

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

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:

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

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

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

которое упрощается для больших отношений сигнал/шум до

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

Для различных типов манипуляции будем иметь:

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

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

$\overline{P_E} \approx 1/2h^2, \quad \gamma_c = \sqrt{2}$ 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

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

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

$$p_{общ} = \prod_{i=1}^{n} p_i = p^n \, . \tag{3.19}$$

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 [7]

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 $\mathrm{Re}\{h_c\}$.

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

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

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

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

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 [7]) 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 временя 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:

$\tau_p = max\ \tau_{зап} – min\ \tau_{зап}$

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 $0 \le t < T$. 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 $s_i(T—t)$, $0 \le t < T$. К выходам фильтров подключены детекторы огибающей, выходные напряжения которых сравниваются в момент времени $t = T$. Индекс i=l, ..., 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.

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

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

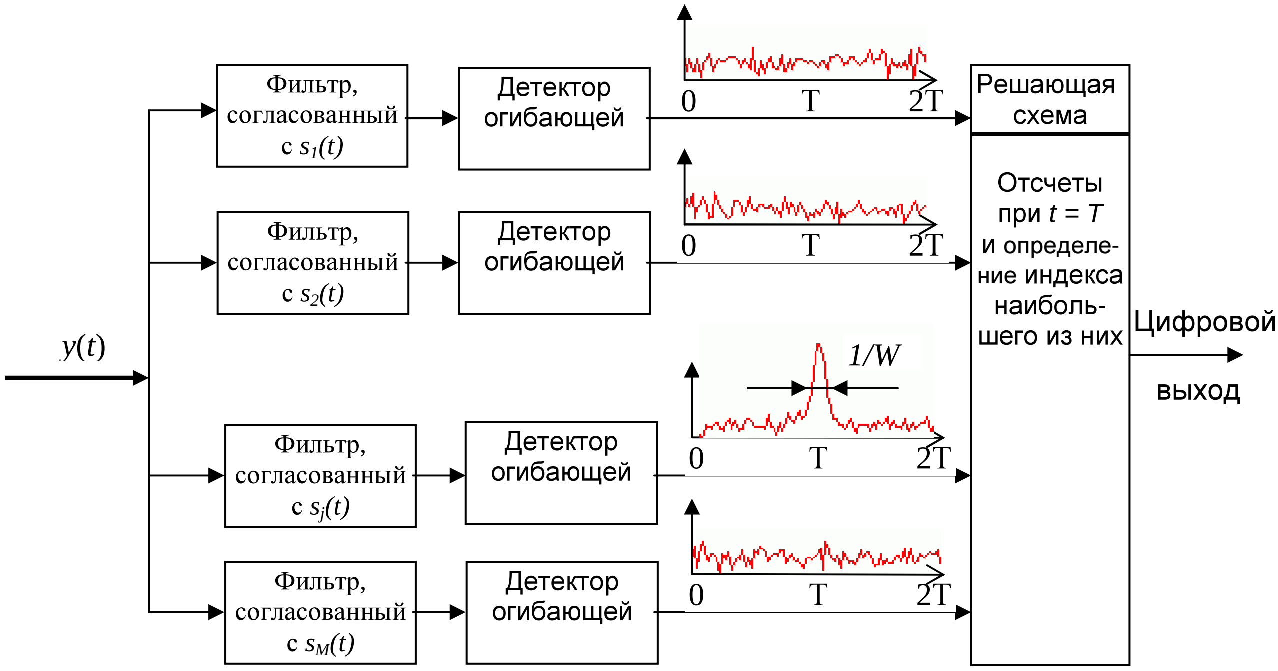


Рис. 3.3. Оптимальный некогерентный приемник для однолучевого канала (равновероятные сигналы равной энергии; гауссовский шум).

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

$$|R_{ik}| << E, \quad k \neq i;$$
$$|R_{ii}| = E, \quad |t| < 1/W. \tag{3.22}$$

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 $0 \le t < T$. Если другой символ, скажем, $s_i\{t)$, был передан сразу после него на временном интервале $T \le t < 2T$, то, очевидно, отклик на него появится на временном интервале $T \le t < 3T$. 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 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: $\Delta = \tau_p$.

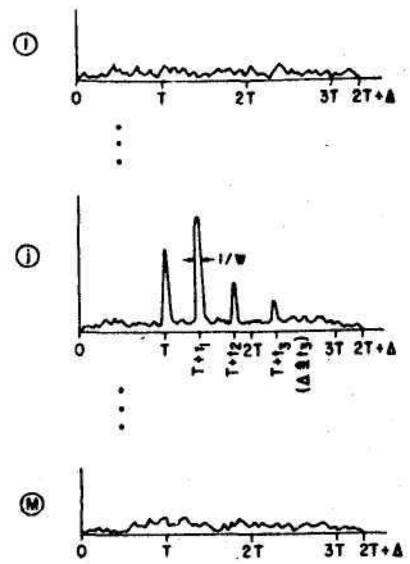


Рис. 3.4. Выходы детекторов огибающих. 4-х лучевой случай.

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

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

$$l = (1, M), \quad 0 \le t < 2T + \Delta. \qquad (3.23)$$

При выполнении условия разрешимости (3.20) главные пики огибающей $e_i(t)$ различны и соответствуют моментам времени t_0=0, t_1, t_2, и t_3 = Δ. 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 $s_j(t)$, $0 \le t < T$.

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 $t_0, \ldots t_{K-1}$, 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 длятся 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, 2л], 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 $T+t_k$, $k=0, \ldots K-1$, 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 l-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 [7]

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

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 [7]:

$$w_l = \sum_{k=0}^{K-1} \frac{\psi_k x_{ik}^2}{N_0 + \psi_k E}, \tag{3.25}$$

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_l 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 ($\psi_k E >> N_0$) 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 $p(t_к)$, $0<t_к<\Delta$, а их индексы переупорядочены в соответствии с ростом величины $t_к$. 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 $W\Delta>>1$.

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

Такой приемник, с учетом сделанных допущений, должен вычислять величины

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

где 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 $h_c(t)$: $h_p(t) = h_c(\Delta\text{-}t)$.

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

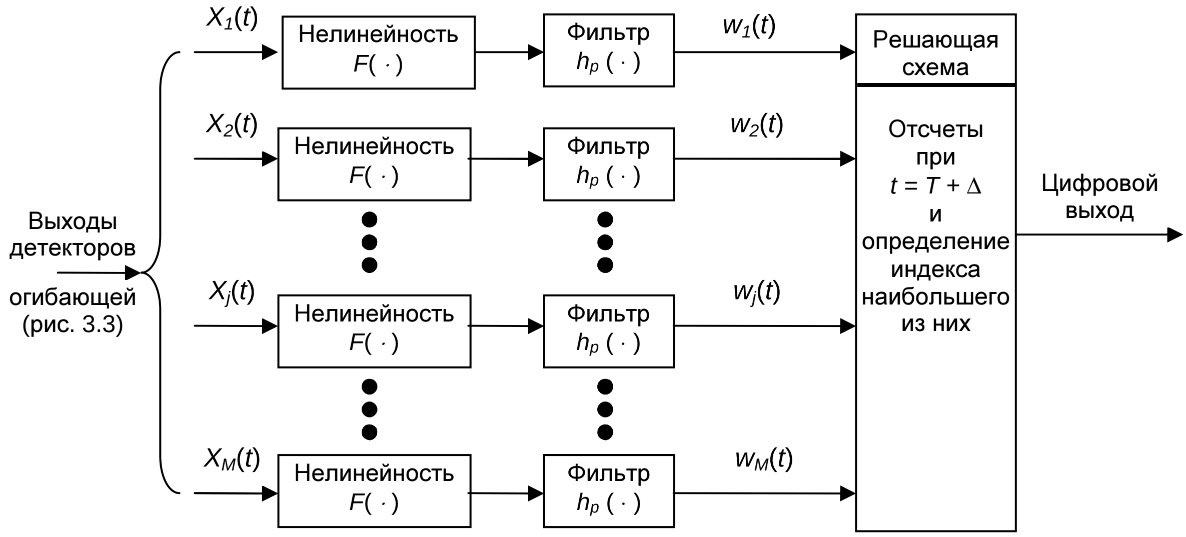


Рис. 3.5. Оптимальный некогерентный приемник для многолучевого канала с неизвестными временами запаздывания лучей.

Следует отметить, что в отличие от выражений (3.24) и (3.25), где отсчеты x_{ik}, $k=0, \ldots, K\text{-}1$, объединяются по линейному и квадратичному законам, для объединения отсчетов в (3,26) используются экстремальные нелинейности экспоненциального типа, показная на рис. 3.6.

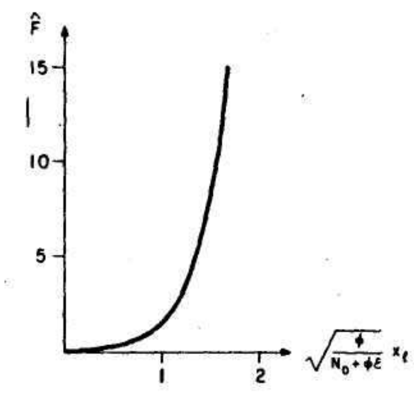


Рис. 3.6. Используемый тип нелинейности $F(\cdot)$

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

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 данном 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).

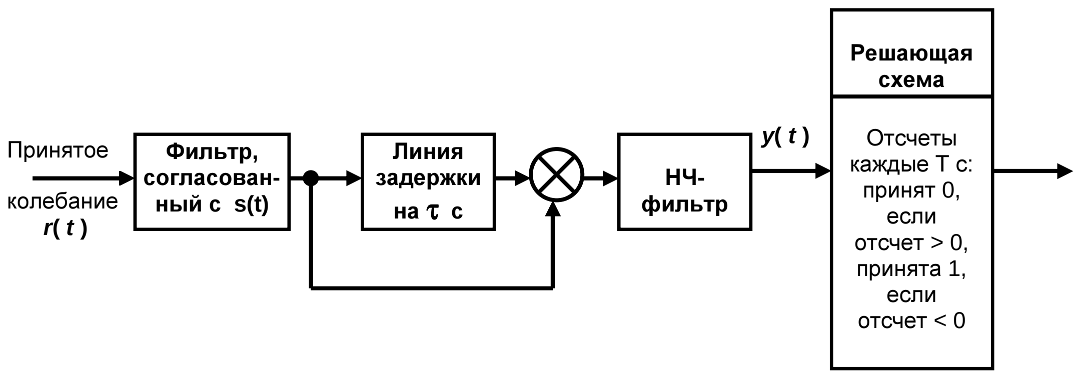


Рис. 3.7. Оптимальный приемник однолучевого ОФМ-сигнала

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.

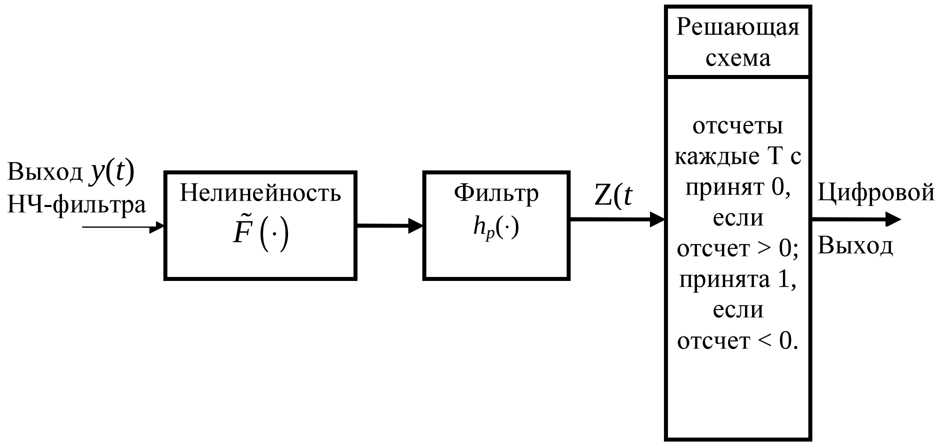
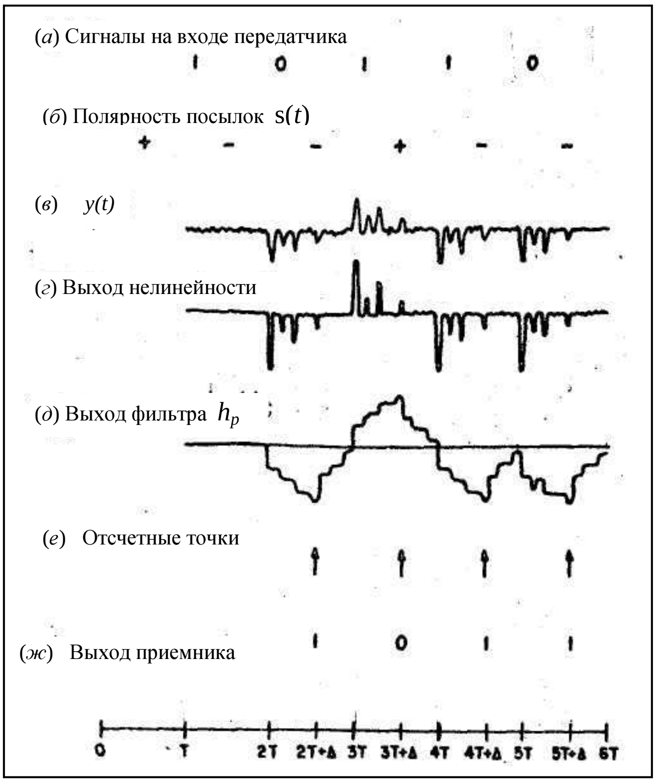


Рис. 3.8. Оптимальный приемник ОФМ-сигналов для многолучевого канала с неизвестными запаздываниями лучей.

На этом рисунке нелинейность

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

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

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

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

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

$$h_p(t) = \begin{cases} 1, & 0 \le t \le \Delta; \\ 0 & \text{при других } t \end{cases} \qquad (3.29а)$$

показаны на рис. 3.9, *г, д*. Отсчеты выхода фильтра $h_p(\cdot)$ берутся в моменты времени $nT+\Delta$, n=2, 3, . . . , и используются для формирования выходных сигналов приемника, задержанных на (T+Δ) 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 специальные сигналы, либо сами информационные сигналы; в последнем случае зондирование и передача данных осуществляются одновременно.

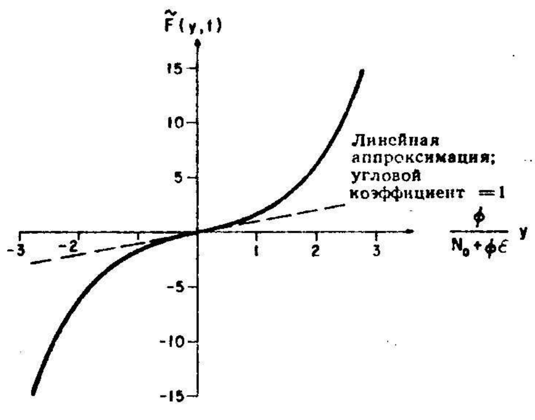


Рис. 3.10. Характеристика нелинейности

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 $h_p(\cdot)$ определять не по априорным, а по апостериорным статистикам. Действительно, такой подход уместен в отношении распределений интенсивностей лучей и нелинейностей F, которые они определяют.

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

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

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

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

так как при больших *x* $1\mathrm{n}I_0$(x) ≈ 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.

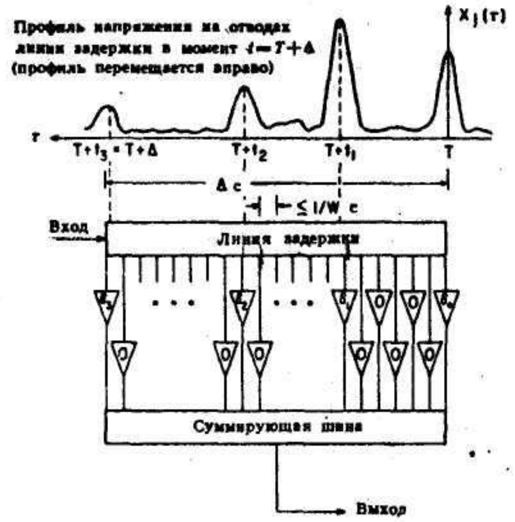


Рис. 3.11. Трансверсальный фильтр.

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 WΔ отводов. На вход трансверсального фильтра подается напряжение с выхода детектора огибающей, например $X_j\{(t)$ (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,.... 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.

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, . . 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=Δ, где Δ — максимальное 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:

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

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 $h_c(t)$.

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

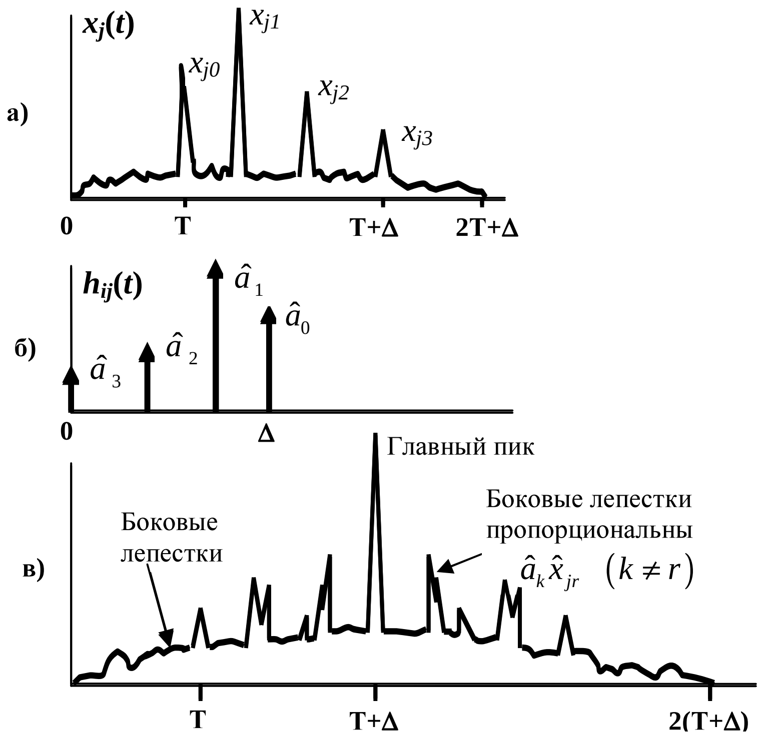


Рис. 3.12. (а) – вход фильтра; (б) – импульсная характеристика фильтра; (в) – Выход фильтра (свертка входа и импульсной характеристики фильтра)

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 лика 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 $x_l(t)$, $l=1,\ldots,M$ 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 большие 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.

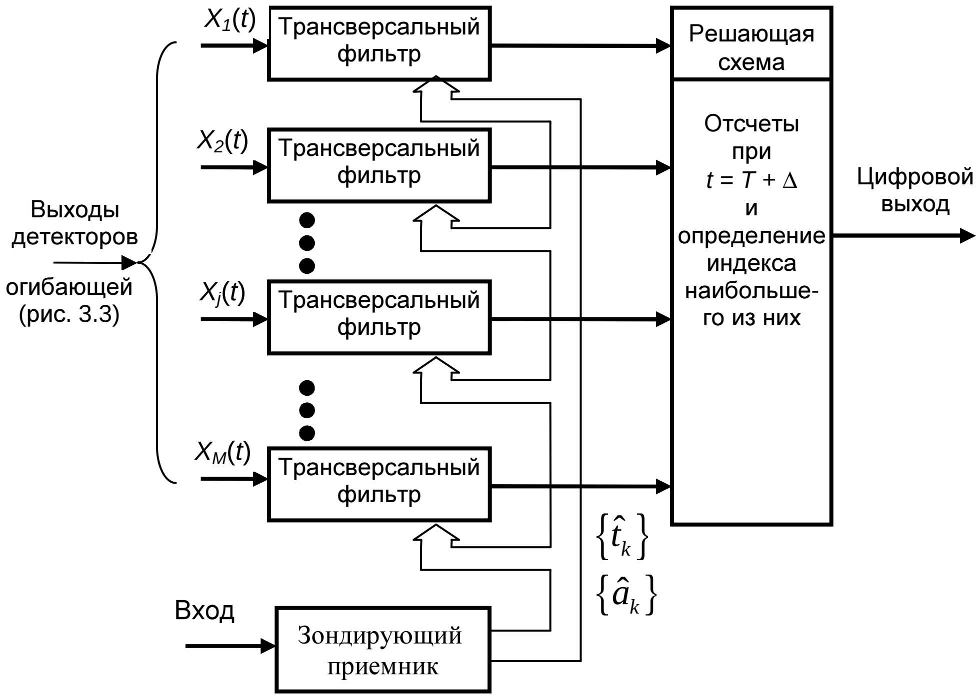


Рис. 3.13. Приемник многолучевого сигнала с зондированием канала (приемник системы «Рейк»).

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 раздельным 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.

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 2Т+Δ, что может привести к межсимвольной интерференции.

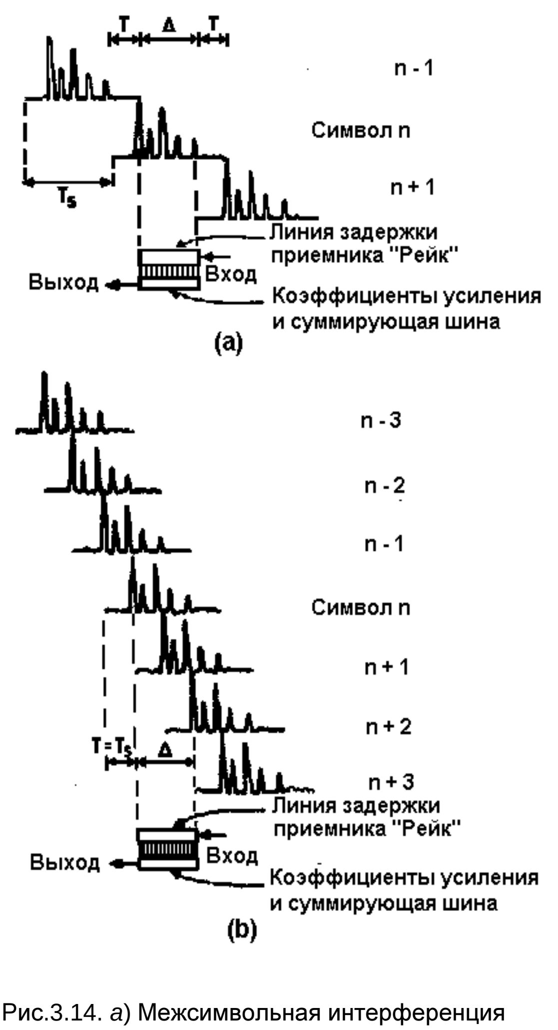


Рис.3.14. *a*) Межсимвольная интерференция отсутствует; *b*) T_S=T; Δ=2T – умеренная межсимвольная интерференция

Для того чтобы наглядно проиллюстрировать возникновение межсимвольной интерференции в широкополосном случае (TW>>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 *a* 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 *b* показан случай умеренной межсимвольной интерференции. Здесь T_s = T, а Δ=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 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[Δ/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*>>1 и ограничений, налагаемых на ширину полосы *W*). Следовательно, в бинарных системах мы не можем полностью устранить межсимвольную интерференцию, если скорость передачи данных превышает величину, равную примерно 1/Δ бит/с. Так, например, при работе такой системы в городских условиях (эффективное значение Δ=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 *b*), 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*>>1 и так как мы теперь полагаем Δ/T>1, то *W*Δ>>1 и это означает, что линия задержки должна иметь большое число отводов. Например, при *TW*=100 и Δ/*T*=5 линия задержки будет иметь *W*Δ=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 будут соответствовать пикам центрального *n*-го отклика на рис. 3.14 *b*; при этом весьма маловероятно, что пики смежных с ним откликов также попадут на эти активированные отводы.

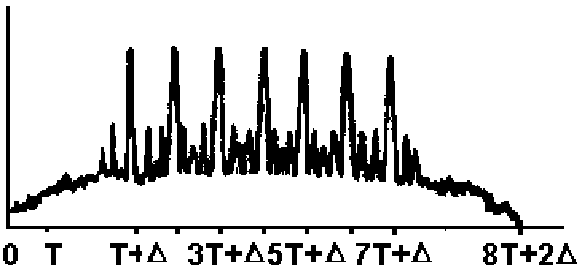


Рис.3.15. Реакция ТФ приемника «Рейк» на 7 сигналов рис. 3.14.

Последовательность импульсов после семи таких сверток и соответствующего их объединения будет иметь примерно такой вид, как показано иа 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/Δ бит/с, скоростями передачи данных без заметного ухудшения характеристик.

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

Приемник с последетекторным интегрированием сигналов

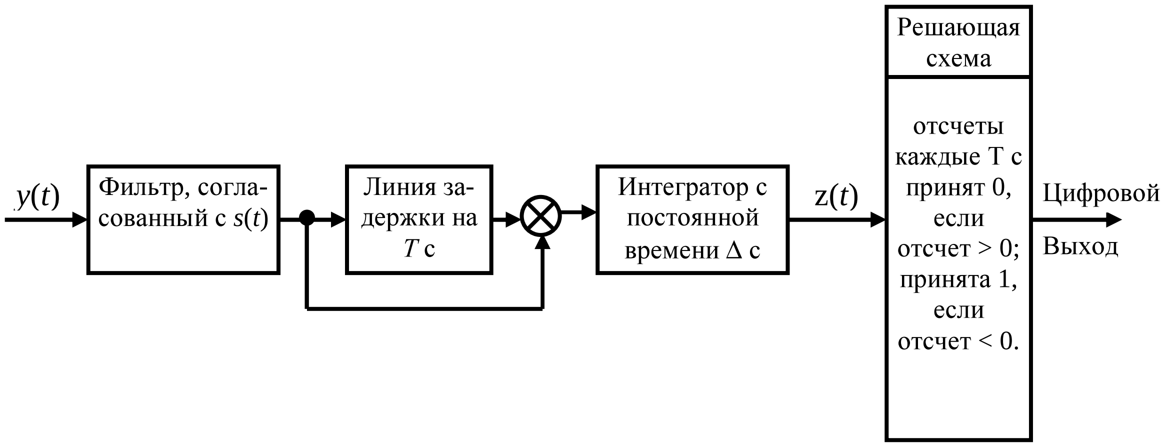


Рис. 3.16. ПДИ – квазиоптимальный приемник ОФМ-сигнала для неизвестных времен запаздывания лучей.

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 (3.29*а*). Эта характеристика предполагает, что запаздывания лучей равномерно распределены на интервале [0, Δ] и $p(t_к)=1/\Delta$, $0 \le t_k \le \Delta$. В действительности запаздывания лучей распределены неравномерно, поэтому характеристика $h_p(t)$, определяемая выражением (3.29*а*), 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 релеевские распределения;
- все лучи разрешимы;
- $\Delta << T$, МСИ отсутствует;
- в канале присутствует аддитивный белый шум с односторонней плотностью мощности N_0.

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

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

Для сравнения на рис. 3.17 указаны значения вероятности ошибки для оптимального однолучевого ОФМ-приемника. Нетрудно видеть, что при наличии ряда релеевских лучей с близкими среднеквадратич- ными 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.

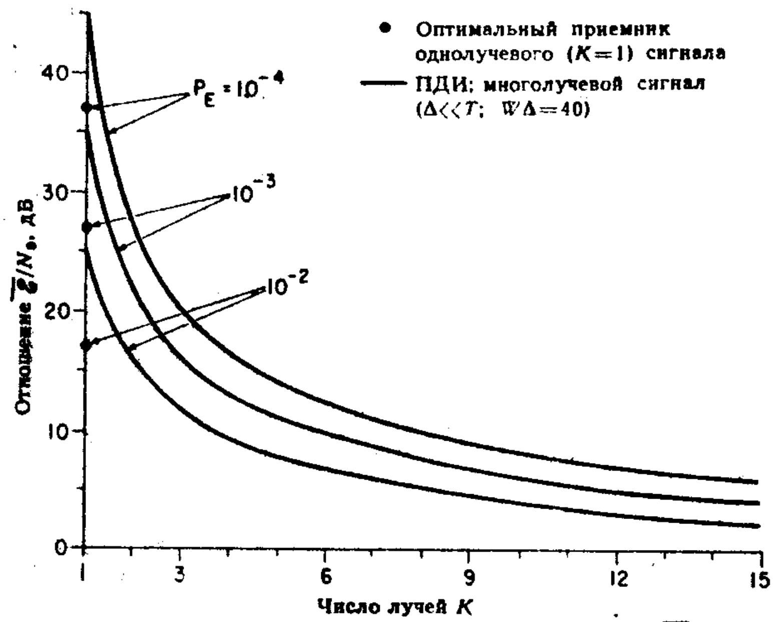


Рис.3.17. **Зависимость среднего на 1 луч отношения $\bar{\mathscr{E}}/N_0$ для ОФМ-приемника с последетекторным интегрированием сигнала, требуемого для обеспечения заданной вероятности ошибки на 1 бит, от числа лучей K (независимые лучи с рэлеевскими замираниями).**

3.2.1.6. 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.

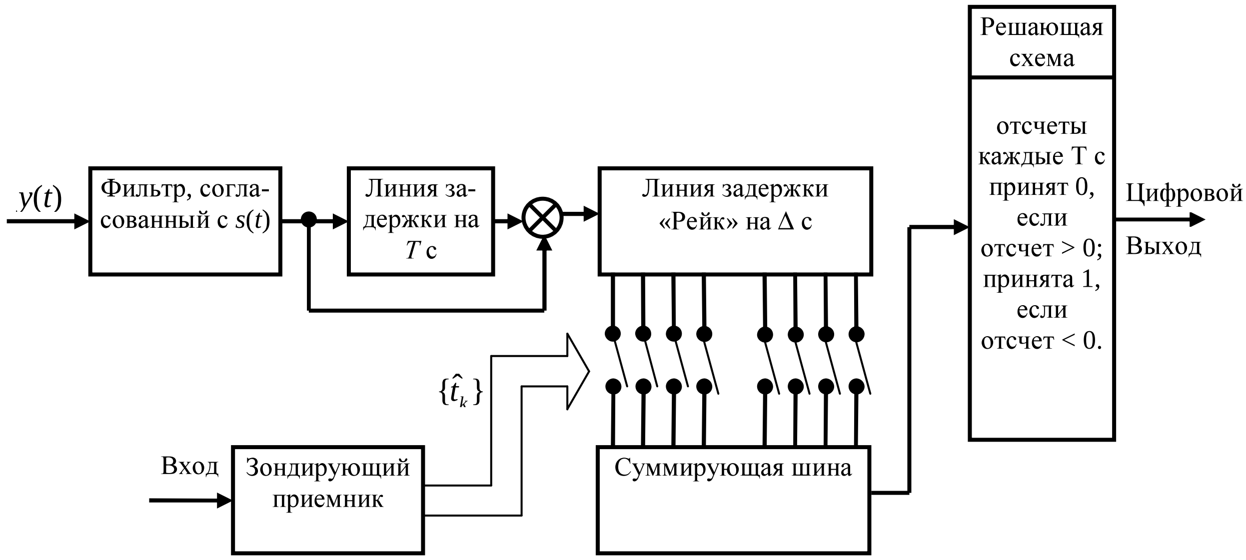


Рис. 3.18. Цифровой вариант приемника «Рейк» – квазиоптимальный приемник ОФМ-сигнала, в котором используются оценки времен запаздывания лучей.

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

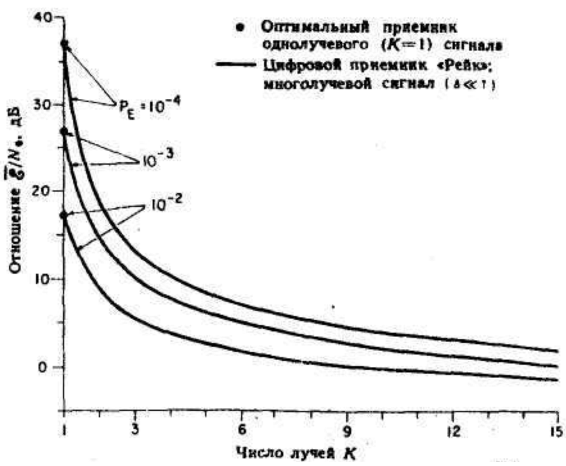


Рис. 3.19. Зависимость среднего на 1 луч ОСШ для цифрового Рейк от числа лучей, требуемого для обеспечения заданной вероятности ошибки на 1 бит.

всех ключей приемник, показанный из рис. 3.18, оказывается примерно аналогичным ФМ-приемнику, показанному на рис. 3.16, у которого операция интегрирования на интервале Δ с заменена операцией суммирования в дискретных точках этого интервала. Характеристики этого приемника, аналогичные кривым рис. 3.17, представлены на рис. 3.19 для случая релеевских замираний без учета межлучевой и межсимвольной интерференции, наличие которой можно учесть введением дополнительного эквивалентного шума.

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, Δ/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 $y(()$.
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 использованным при определении первого луча (ПЛ).
5) *Взвешенный последетекторный интегратор (ВПДИ).* Установив тот факт, что последний луч в профиле многолучево-

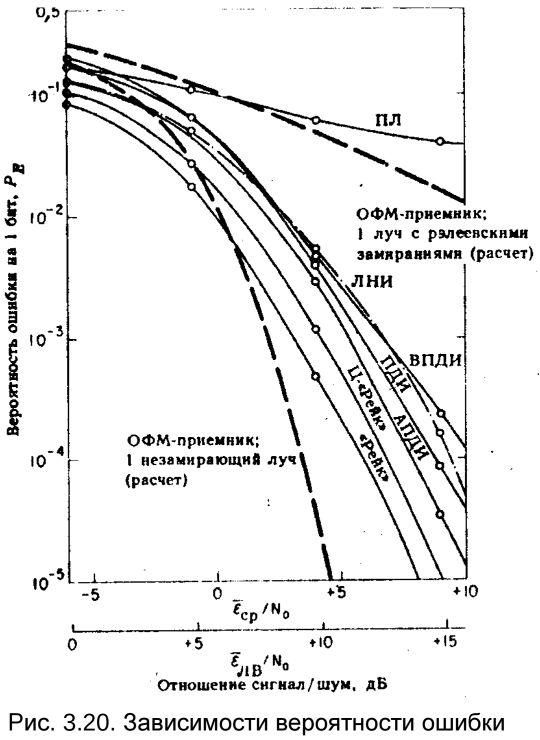


Рис. 3.20. Зависимости вероятности ошибки на 1 бит от величины ОСШ для различных приемников в условиях многолучевого распространения радиоволн в районе с плотной многоэтажной застройкой [xxx]/

сти является, как правило, более слабым и менее частым, чем предыдущие лучи, прямоугольную импульсную характеристику $h_p(t)$, определяемую выражением (3.29*а*), заменили экспоненциальной весовой функцией $h_p(t)=\exp[-(\Delta-t)/\tau]$, $0\le t\le T$, 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 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 I 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 [8], 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 [8].

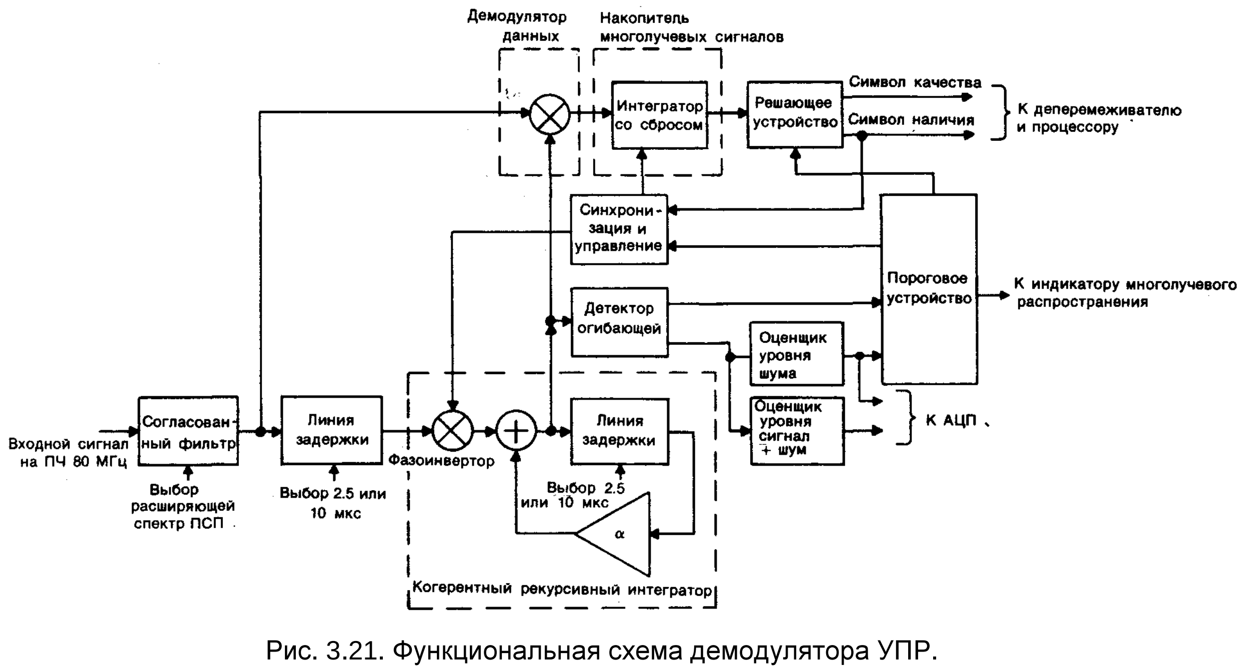


Рис. 3.21. Функциональная схема демодулятора УПР.

Не касаясь других, подчас довольно интересных, технических решений и характеристик указанной УПР, детально описанных в [8], рассмотрим демодулятор этой установки (рис. 3.21).

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, 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 длитель-

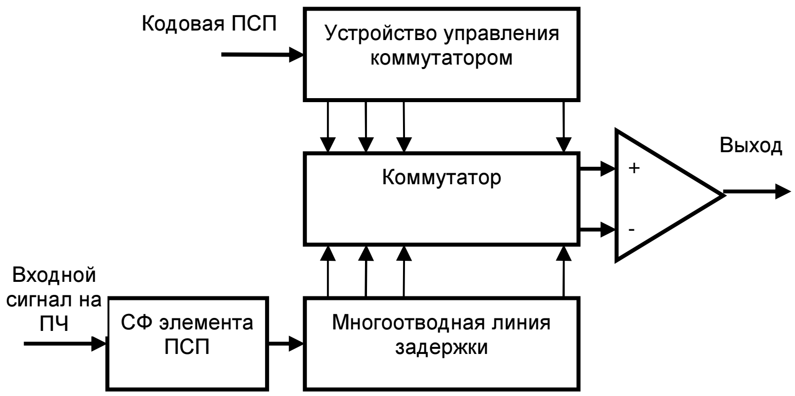


Рис. 3.22. Программируемый согласованный фильтр

ность 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 а контура КРИ (рис. 3) и равно $(1+\alpha)/(1-\alpha)$. При α==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 распространении 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 ~ 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 удается избегать.

Список литературы

1. Скляр 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. 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.
6. Файфер У.С., Бруно Ф.Дж. Недорогая приемопередающая установка пакетной радиосвязи. ТИИЭР, т. 75, № 1, январь 1987. С. 41 – 53.