

А. Б. Сергиенко

ЦИФРОВАЯ СВЯЗЬ

Санкт-Петербург
2012

МИНОБРНАУКИ РОССИИ

Санкт-Петербургский государственный
электротехнический университет «ЛЭТИ»

А. Б. Сергиенко

ЦИФРОВАЯ СВЯЗЬ

УЧЕБНОЕ ПОСОБИЕ

Рекомендовано

*учебно-методическим объединением вузов Российской Федерации по образованию
в области радиотехники, электроники, биомедицинской техники и автоматизации
в качестве учебного пособия для студентов высших учебных заведений, обучающихся
по направлению 210400 «Радиотехника»*

Санкт-Петербург
Издательство СПбГЭТУ «ЛЭТИ»
2012

УДК 621.391
ББК 3 88я7
С32

Сергиенко А. Б. Цифровая связь: Учеб. пособие. СПб.: Изд-во С32 СПбГЭТУ «ЛЭТИ», 2012. 164 с.

ISBN 978-5-7629-1202-0

Рассматриваются основные принципы функционирования систем цифровой связи, используемые в них алгоритмы обработки сигнала, способы оценки помехоустойчивости таких систем.

Предназначено для студентов факультета радиотехники и телекоммуникаций направления 210400 «Радиотехника».

УДК 621.391
ББК 3 88я7

Рецензенты: кафедра цифровой обработки сигналов Санкт-Петербургского государственного университета телекоммуникаций им. проф. М. А. Бонч-Бруевича;
доц. В. А. Варгаузин (Санкт-Петербургский государственный политехнический университет).

Утверждено
редакционно-издательским советом университета
в качестве учебного пособия

ISBN 978-5-7629-1202-0

© СПбГЭТУ «ЛЭТИ», 2012

1. ОБЩАЯ СТРУКТУРА И ОСНОВНЫЕ ХАРАКТЕРИСТИКИ СИСТЕМ ЦИФРОВОЙ СВЯЗИ

1.1. Понятие цифровой связи и ее краткая история

Под термином «цифровая связь» понимают передачу сообщений, которые представляют собой элементы *конечного* множества — *символы*.

Передача сообщений такого рода используется человеком с древнейших времен. Первобытные люди передавали *сигналы*, используя дым костров и звуки ударов. Письменность — также передача сообщений цифрового типа. Более того, можно сказать, что алфавитное письмо за счет присущей ему избыточности (она помогает нам правильно воспринимать тексты с некоторым количеством опечаток) обладает элементами *помехоустойчивого кодирования*.

Передача информации при помощи огней и других оптических сигналов применялась в течение многих веков (вспомним хотя бы морской флажковый семафор). В 1793 г. французский инженер Клод Шапп создал *оптический телеграф*, в котором появились некоторые черты *сетей связи* — сообщение передавалось по цепочке *ретрансляторов*, представлявших собой башни с установленным на них механизмом из рычагов. Разным символам соответствовало разное положение рычагов, что позволяло передать сообщение на следующий ретранслятор, расположенный в пределах прямой видимости. Самая длинная линия такого телеграфа (Санкт-Петербург — Варшава, 1839 г.) имела протяженность 1200 км, содержала 149 башен-ретрансляторов и позволяла передать сообщение с одного конца линии на другой за 22 мин.

Начало использования электричества для целей цифровой связи относится к 1830-м гг., когда появились несколько разновидностей электрического телеграфа (П. Л. Шиллинг — 1832 г., С. Морзе — 1837 г.). Аппарат Морзе и созданный Морзе телеграфный код получили наибольшее распространение и использовались в течение многих десятилетий.

В 1870-е гг. французский инженер Жан Морис Эмиль Бодо создал телеграфный аппарат, в котором для передачи символов использовался 5-элементный код (5-битовый, как сказали бы сейчас), т. е. впервые применены *многобитовые символы фиксированного размера*. Разработанный Бодо код с некоторыми усовершенствованиями применялся в телеграфной технике вплоть до конца XX в.

Изобретение А. С. Поповым радиосвязи (1895 г.) позволило реализовать телеграфную связь с использованием радиоканала.

Новый толчок развитие теории и техники цифровой связи получает после Второй мировой войны. Так, в конце 1940-х гг. создается *теория потенциальной помехоустойчивости* (В. А. Котельников, К. Шеннон).

В значительной мере развитие цифровой связи стимулировалось появлением вычислительной техники. Были необходимы средства обмена инфор-

мацией между компьютерами, в результате в 1950-е гг. появились первые *модемы*, использовавшие в качестве каналов связи готовую инфраструктуру телефонных линий. Несколько позже, по мере развития теории и техники *цифровой обработки сигналов*, начался постепенный переход телефонных сетей на цифровую основу.

К концу XX в. сформировалось новое основное направление развития цифровой связи — сети *мобильной связи*, сначала существовавшие в виде отдельных систем *сотовой телефонии* и систем *беспроводной передачи данных*. Наконец, в начале XXI в. формируется тенденция слияния (*конвергенции*) этих систем с целью дать пользователю возможность обмена сообщениями любого типа в любой точке планеты.

1.2. Общая структура системы цифровой связи

Назначением системы связи является передача *сообщения* из одной точки в другую через *канал связи*, обладающий определенными свойствами (в частности, пропускающий лишь некоторую полосу частот). Для решения этой задачи приходится осуществлять целый ряд преобразований. Общая структура системы цифровой связи показана на рис. 1.1.

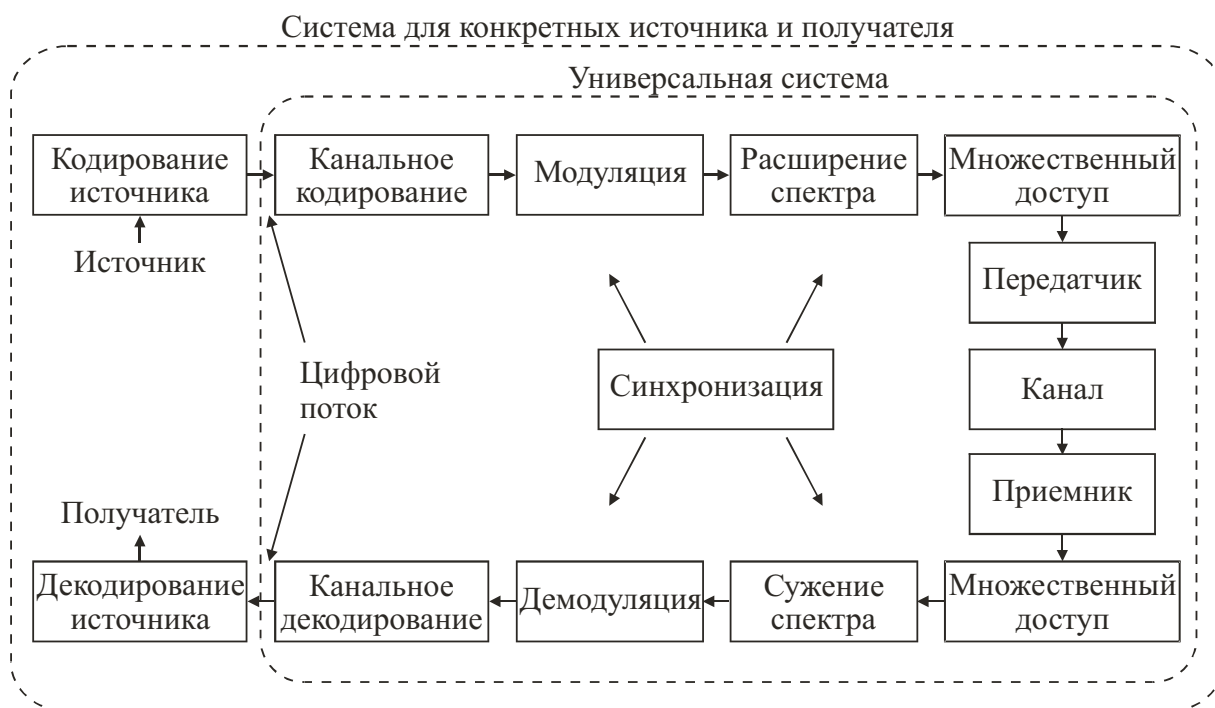


Рис. 1.1. Блок-схема системы цифровой связи

Прежде всего исходное сообщение подвергается первичному кодированию (*кодированию источника*), цель которого — преобразование аналогового сообщения в цифровое либо сжатие информации.

Следующий этап — *канальное кодирование*. Целью канального кодирования является учет таких свойств канала связи, как уровень шума и измен-

чивость во времени. Чаще всего под термином «канальное кодирование» понимается *помехоустойчивое кодирование*. При этом в сообщение вносится избыточность с целью обеспечить возможность исправления на приемной стороне всех или некоторых возникших в процессе передачи ошибок. Также к канальному кодированию относятся технологии перемежения, используемые для борьбы с замираниями сигнала.

После применения помехоустойчивого кода сообщение поступает в *модулятор*, преобразующий цифровое сообщение в аналоговый модулированный сигнал, занимающий заданную полосу частот.

Следующие 2 блока являются необязательными. Дополнительное *расширение спектра* позволяет получить ряд преимуществ. Технологии *множественного доступа* позволяют реализовать совместное использование общего частотно-временного ресурса несколькими радиолиниями.

Далее сигнал поступает в *передатчик* — аналоговый блок, осуществляющий усиление сигнала и реализующий интерфейс с каналом связи (антенная система для радиоканала, согласующие цепи для проводного канала и т. п.).

В процессе прохождения модулированного сигнала через *канал связи* сигнал подвергается воздействию шумов и помех. Искаженный сигнал поступает на вход приемника.

Структура приемной части является зеркальным отражением структуры передатчика — сигнал проходит через блоки, в обратном порядке осуществляющие преобразования, обратные по отношению к тем, что производились в передатчике.

Приемник реализует интерфейс с каналом связи и осуществляет предварительную обработку аналогового сигнала. Далее (при необходимости) происходит выделение нужного сигнала в соответствии с используемым методом *множественного доступа* и выполняется снятие введенного ранее *расширения спектра*.

Потом сигнал подвергается *демодуляции*, в процессе которой аналоговый модулированный сигнал преобразуется в цифровое сообщение. Далее производится канальное декодирование (*декодирование помехоустойчивого кода*), при этом благодаря корректирующим свойствам кода возможно исправление части (или всех) ошибок, возникших в процессе передачи. После исправления ошибок следует *декодирование источника* — восстановление исходного сообщения.

1.3. Специализированные и универсальные системы связи

Специализированная система ориентирована на конкретный источник информации, она содержит кодер и декодер источника, адаптированные к его

свойствам. Пример — сотовый телефон, предназначенный в первую очередь для передачи речевого сигнала.

Универсальная система получает на входе цифровой битовый поток и ничего не знает о его статистических свойствах. Ее задача — донести этот поток до получателя.

В настоящее время все больше систем становятся универсальными. Иными словами, кодер источника все чаще оказывается отделен от собственно системы связи (хотя совместная оптимизация системы, затрагивающая несколько ее уровней одновременно, позволяет достичь лучших результатов).

По этой причине далее рассматриваются только универсальные системы связи.

1.4. Важнейшие параметры систем цифровой связи

Системы цифровой связи характеризуются рядом технических параметров, однако прежде чем перейти к их описанию, необходимо обсудить смысл двух важных терминов: *бит* и *символ*.

- *Бит* — единица количества информации, но в то же время это элемент входного потока данных, который может принимать значения 0 и 1. Так как рассматриваются универсальные системы связи, которые ничего не знают о свойствах поступающего для передачи цифрового потока, предполагается, что поступающие для передачи биты равновероятно и независимо друг от друга принимают значения 0 и 1. Чтобы обеспечить это при любом характере поступающих на вход данных, в реальных системах используется *скремблирование* (см. 1.5).
- *Символ* — практически во всех системах связи на тех или иных этапах обработки информации биты группируются для одновременной передачи (пример — параллельный 8-битовый порт персонального компьютера). Такая группа бит называется *символом*. Число бит в символе обозначается буквой m , число возможных значений символа — буквой M , так что $M = 2^m$. Если символ состоит из одного бита ($m = 1$, $M = 2$), такой способ передачи называют *бинарным*. Преобразование последовательного битового потока в m -битовые символы и обратно неизбежно вносит задержку в процесс передачи данных. Период, с которым символы следуют друг за другом, далее будет обозначаться буквой T . Вследствие указанных выше предположений о свойствах битового потока передаваемые символы также предполагаются статистически независимыми и равновероятно принимающими все M возможных значений.
- *Скорость передачи данных*:
 - *битовая* — число бит, переданных в единицу времени — R , бит/с;

- *символьная* — число символов, переданных в единицу времени — $R_{\text{sym}} = R/m$, симв/с. Эта величина обратна длительности передачи символа: $R_{\text{sym}} = 1/T$.
- *Ширина спектра сигнала* — W , Гц. Теоретически для любого сигнала конечной длительности ширина спектра бесконечна, поэтому здесь подразумевается измерение ширины по какому-либо практическому критерию.
- *Спектральная эффективность* — R/W , (бит/с)/Гц, показывающая соотношение между битовой скоростью передачи данных и шириной спектра сигнала. Данный параметр определяет, насколько эффективно система использует полосу частот для передачи информации.
- *Вероятность ошибки*, т. е. вероятность того, что принятое числовое значение не совпадает с переданным:
 - *битовая* — вероятность того, что принятое значение *бита* не совпадает с переданным, обозначается как P_b или BER (Bit Error Rate), $P_b \in [0, 0,5]$;
 - *символьная* — вероятность того, что принятое значение *символа* не совпадает с переданным, обозначается как P_{sym} или SER (Symbol Error Rate), $P_{\text{sym}} \in [0, (M - 1)/M]$.

Замечание. Строго говоря, BER и SER — это не *вероятность* ошибок, а *частота* их появления во времени, т. е. доля от общего числа переданных бит или символов. Однако, если случайный процесс возникновения ошибок является эргодическим, эти величины равны соответствующим вероятностям ($\text{BER} = P_b$, $\text{SER} = P_{\text{sym}}$).

Поскольку символ состоит из нескольких бит, в случае ошибки в нем, скорее всего, искажаются не все биты, а лишь какая-то их часть. Поэтому вероятность битовой ошибки никогда не превышает вероятности символьной ошибки: $P_b \leq P_{\text{sym}}$ (равенство наблюдается в бинарных системах, где понятия бита и символа совпадают: $P_b = P_{\text{sym}}$).

Указанные верхние границы для P_b и P_{sym} вытекают из следующих соображений: вероятность битовой ошибки не может превышать 50 %, так как именно в этом случае будет наблюдаться полное отсутствие статистической связи между переданным и принятым сообщениями (при этом с вероятностью 50 % значение принятого бита просто угадывается). Если вероятность битовой ошибки превышает 50 %, всегда можно инвертировать полученный результат, сделав тем самым вероятность ошибки меньше 50 %. Что касается предельной вероятности символьной ошибки, то при отсутствии статистической связи между переданным и принятым сообщениями вероятность случайного угадывания правильного символа составляет $1/M$, а следовательно, вероятность символьной ошибки равна $(M - 1)/M$.

- *Отношение сигнал/шум* в системах цифровой связи может измеряться по-разному:

- *отношение сигнал/шум по мощности* — SNR, дБ. Показывает отношение мощности сигнала к мощности шума. При использовании модели белого шума его полная мощность является бесконечной, поэтому для использования данного параметра необходимо определить, в какой полосе частот будет измеряться мощность шума. Как правило, эта полоса частот совпадает с полосой частот полезного сигнала. Таким образом, данный параметр математически определяется следующим образом:

$$\text{SNR} = \frac{P_s}{P_n} = \frac{P_s}{N_0 W}, \quad (1.1)$$

где P_s — мощность сигнала; N_0 — спектральная плотность мощности (СПМ) белого шума; W — ширина спектра сигнала.

Данный параметр является привычной мерой для измерения отношения сигнала к шуму в аналоговых системах, однако для цифровых систем связи он недостаточно информативен. Для сравнения эффективности разных схем модуляции и кодирования сигнала применяется другая мера отношения сигнал/шум [1, 3.1.4]:

- *отношение сигнал/шум на бит*, определяемое как отношение энергии сигнала E_b , приходящейся на один бит передаваемых данных, к СПМ белого шума — E_b/N_0 , дБ.

Для установления связи между отношениями сигнал/шум, вычисляемыми двумя указанными способами, необходимо заметить, что мощность сигнала равна энергии бита, умноженной на битовую скорость: $P_s = E_b R$. С учетом этого из (1.1) получаем

$$\text{SNR} = \frac{E_b R}{N_0 W} = \frac{E_b}{N_0} \frac{R}{W}, \quad (1.2)$$

т. е. SNR равно произведению E_b/N_0 и спектральной эффективности системы.

При использовании данной меры отношения сигнал/шум устраняется зависимость, во-первых, от способа кодирования и модуляции сигнала (за счет того, что рассматривается энергия одного бита передаваемых данных) и, во-вторых, от ширины спектра сигнала (так как в расчете участвует не мощность шума, а его СПМ).

Аналогичным образом можно ввести отношение сигнал/шум *на символ*:

$$\frac{E_{\text{sym}}}{N_0} = \frac{m E_b}{N_0}.$$

1.5. Скремблирование

Целью скремблирования передаваемых данных является придание битовому потоку свойств псевдослучайной последовательности и снижение вероятности появления в нем длинных последовательностей нулей или единиц, а также периодических фрагментов. Сущность скремблирования состоит в контролируемом и легко обратимом изменении значений бит передаваемого двоичного потока. В системах связи получили распространение скремблеры двух типов — аддитивные и мультипликативные, различающиеся способом формирования скремблированного сигнала и по-разному влияющие на помехоустойчивость.

Структура *аддитивного* скремблера и соответствующего ему дескремблера показана на рис. 1.2. Сущность работы скремблера состоит в том, что передаваемый битовый поток *суммируется* (по модулю два) с псевдослучайной последовательностью, генерируемой с помощью регистра сдвига (рис. 1.2, а). В приемнике необходимо произвести такое же преобразование еще раз (рис. 1.2, б).

Структура обратных связей в регистре выбирается таким образом, чтобы он генерировал последовательность максимальной длины (*M*-последовательность). С помощью *n*-битового регистра при должной конфигурации обратных связей можно генерировать псевдослучайную последовательность с периодом $2^n - 1$ бит [2, 8.1.3], [3, 6.6.3].

Чтобы после дескремблера была корректно восстановлена передаваемая информация, регистры сдвига на передающей и приемной сторонах должны работать синхронно. По этой причине данный тип скремблирования, как правило, используется в системах связи, работающих в пакетном режиме. При этом начальное состояние сдвигового регистра, которое было использовано при скремблировании данных, может помещаться в служебную часть (заголовок) пакета, определяться на основании системного времени или являться одинаковым для всех пакетов. Так работают, в частности, системы Bluetooth и Wi-Fi. В этих системах используется 7-битовый регистр, показанный на рис. 1.2.

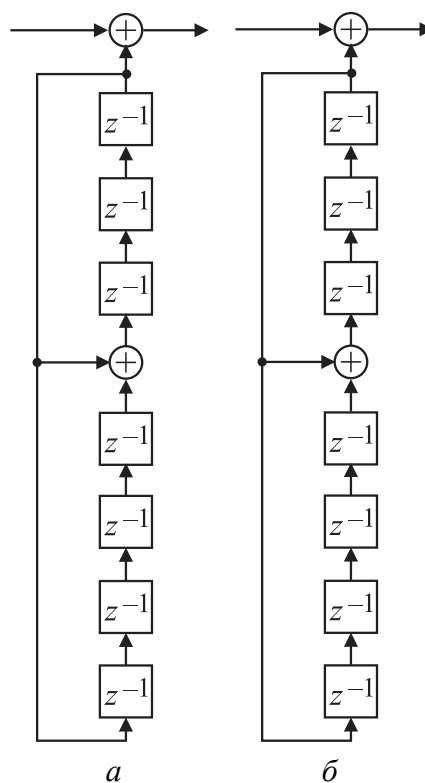


Рис. 1.2. Аддитивный скремблер (а) и соответствующий ему дескремблер (б)

Достоинством аддитивного скремблирования является то, что оно не ухудшает помехоустойчивость системы (разумеется, при правильной синхронизации на приемной стороне).

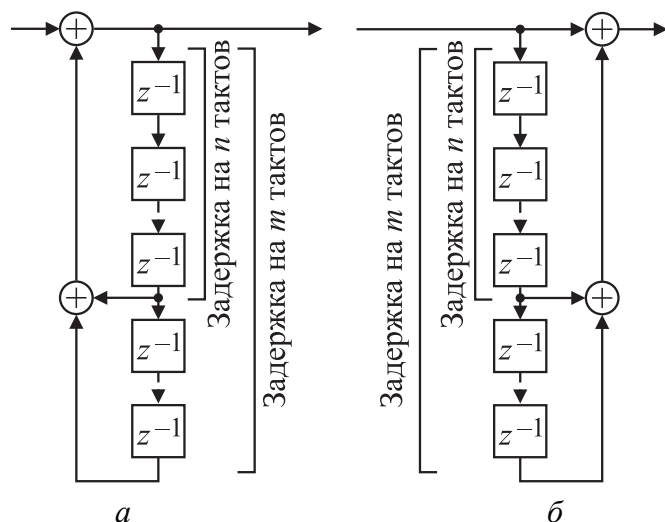


Рис. 1.3. Мультипликативный скремблер (а) и соответствующий ему дескремблер (б)

При мультипликативном скремблировании передаваемая битовая последовательность *делится* на скремблирующий полином, т. е. схема скремблера представляет собой рекурсивный фильтр, вычисления в котором производятся по модулю два (рис. 1.3, а). Конфигурация обратных связей в этом фильтре должна обеспечивать генерирование M -последовательности при нулевом входном сигнале.

Начальное внутреннее состояние фильтра задается при включении передатчика — в регистры памяти записываются заранее оговоренные значения.

Для восстановления передаваемой информации на приемной стороне принятая битовая последовательность *умножается* на скремблирующий полином, что соответствует нерекурсивному фильтру с вычислениями по модулю два (рис. 1.3, б).

Так как на приемной стороне используется нерекурсивная структура, мультипликативный дескремблер не требует синхронизации (поэтому в ряде источников данный способ скремблирования называется самосинхронизирующимся). По этой причине мультипликативное скремблирование применяется в системах, ориентированных на непрерывную передачу данных (модемы, системы xDSL и т. п.).

Однако такой подход приводит к размножению ошибок — из рис. 1.3 видно, что одиночная ошибка в битовом потоке на входе дескремблера приведет к появлению на его выходе нескольких ошибок, число которых равно числу ненулевых коэффициентов скремблирующего полинома. Поэтому для реализации мультипликативного скремблирования используют полиномы с минимально возможным (равным трем) числом ненулевых коэффициентов. Так, во многих протоколах связи модемов голосовой полосы частот и xDSL-модемов используется мультипликативное скремблирование на основе зеркально-симметричных полиномов (для двух направлений связи используются 2 разных полинома) следующего вида:

$$x^{23} + x^5 + 1 \quad \text{и} \quad x^{23} + x^{18} + 1.$$

В данном случае дескремблер повышает вероятность битовой ошибки примерно в 3 раза. Полиномы с тремя ненулевыми коэффициентами существуют не для всех степеней; в частности, они отсутствуют для степеней 8, 12, 13, 14, 16,

1.6. Алгоритм Витерби

При решении ряда задач в системах цифровой связи возникает необходимость поиска оптимальной последовательности бит, символов или их групп, которые могут следовать друг за другом не в любом порядке. Такие задачи могут быть представлены как поиск кратчайшего пути по регулярной решетке, к ним, в частности, относятся:

- прием сигналов с памятью (см. 2.2.7);
- прием сигналов при наличии межсимвольной интерференции (см. 3.2);
- декодирование сверточных кодов и сигнально-кодовых конструкций (см. 8.3.2, 8.4.1).

Оптимальные решения перечисленных задач могут быть найдены с помощью *алгоритма Витерби*. Рассмотрим его сущность на обобщенном примере. На рис. 1.4 приведена схема расположения городов «А» ... «Т» с указанием возможных дорог между ними и их длин. Необходимо найти кратчайший маршрут от «А» до «Т» и узнать его длину.

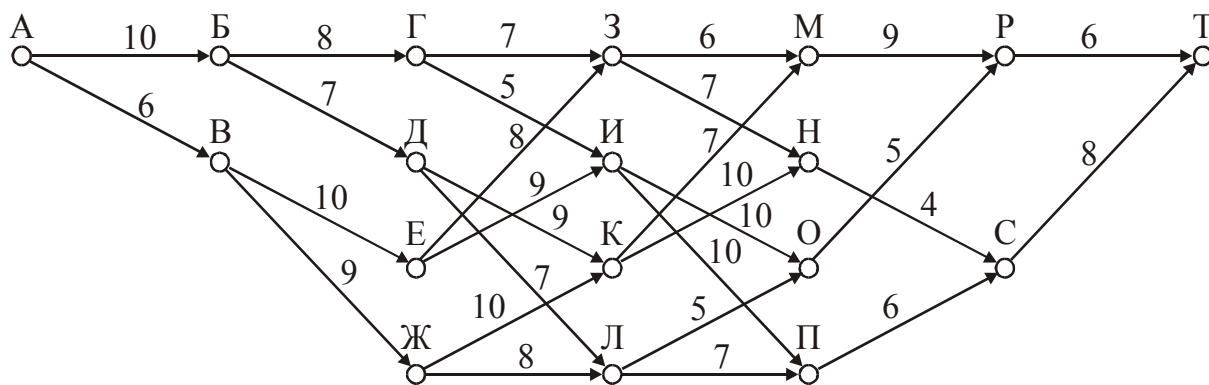


Рис. 1.4. Исходная решетка расстояний

Алгоритм Витерби — это «умный» алгоритм полного перебора, в котором сразу же отбрасываются заведомо бесперспективные пути.

Начинаем продвигаться по решетке слева направо, отслеживая *все* возможные пути и их длины.

На первых двух этапах получаем 4 пути до городов «Г» ... «Ж», при этом пока что не появляется никаких альтернативных вариантов (рис. 1.5).

При переходе к третьему этапу (столбец с городами «З» ... «Л») ситуация меняется. Видно, что к каждому из них ведут 2 пути от городов предыдущей группы. Эти пути имеют разные длины, и можно заметить, что частью

итогового оптимального (кратчайшего) пути может стать *только более короткий* из двух путей, ведущий в каждый из городов.

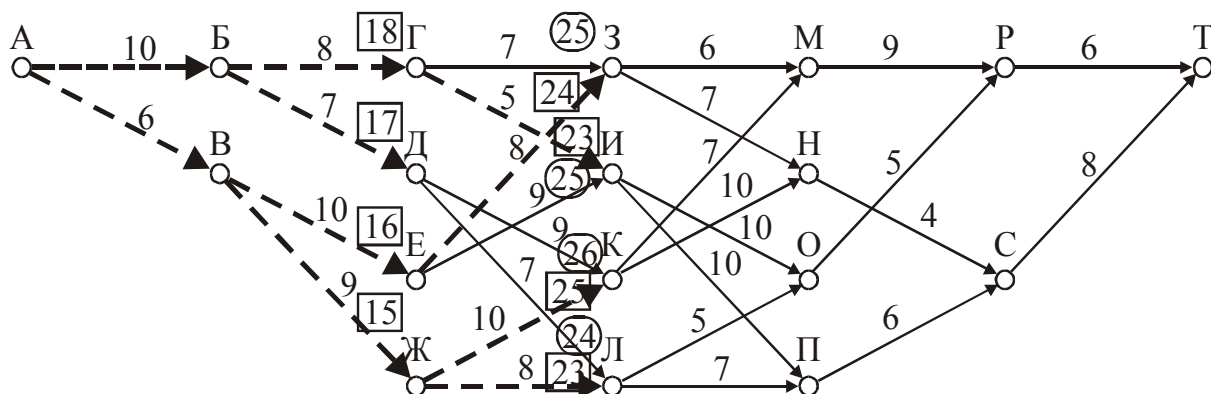


Рис. 1.5. Первый выбор альтернативных вариантов (этап 3)

Доказательство: предположим, что это не так, и частью оптимального пути стал путь, проходящий, например, через город «З» и входящий слева в него не по нижней (А-В-Е-З), а по верхней стрелке (А-Б-Г-З). Но в таком случае всегда можно заменить в итоговом пути фрагмент «АБГЗ» на фрагмент «АВЕЗ», сократив тем самым путь на 1 км, что противоречит предположению об оптимальности итогового пути.

Таким образом, необходимо выбрать для каждого города *более короткий* из входящих в него слева путей. На рис. 1.5 подписаны все варианты расстояний, выбранные пути показаны штриховыми стрелками, а соответствующие им расстояния (*метрики путей*) помещены в квадратные рамки (длины более длинных путей обведены кружочками). Выбранные пути называются *выжившими*.

На следующем этапе (города «М» ... «П») процедура выбора выживающих путей повторяется (рис. 1.6).

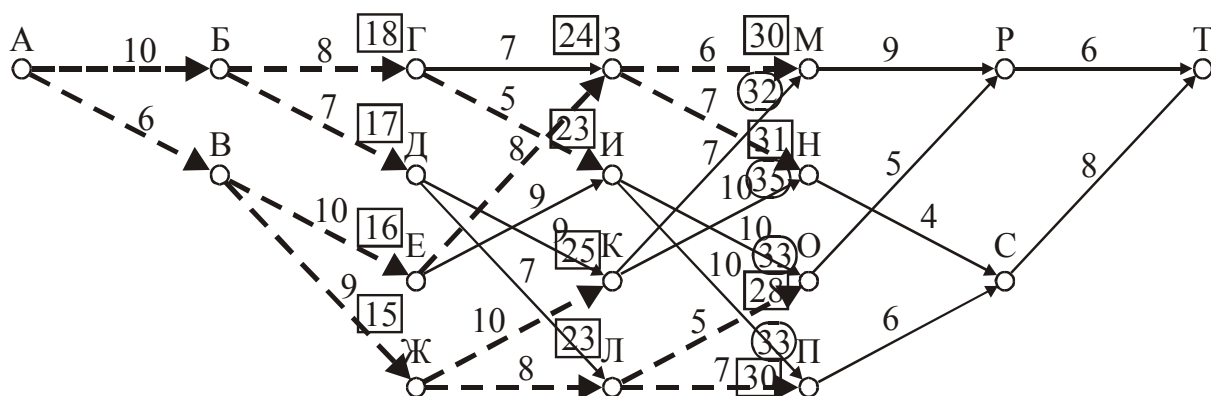


Рис. 1.6. Второй выбор альтернативных вариантов (этап 4)

На следующем этапе смысл действий остается прежним, но число городов сокращается — теперь их только два: «Р» и «С» (рис. 1.7).

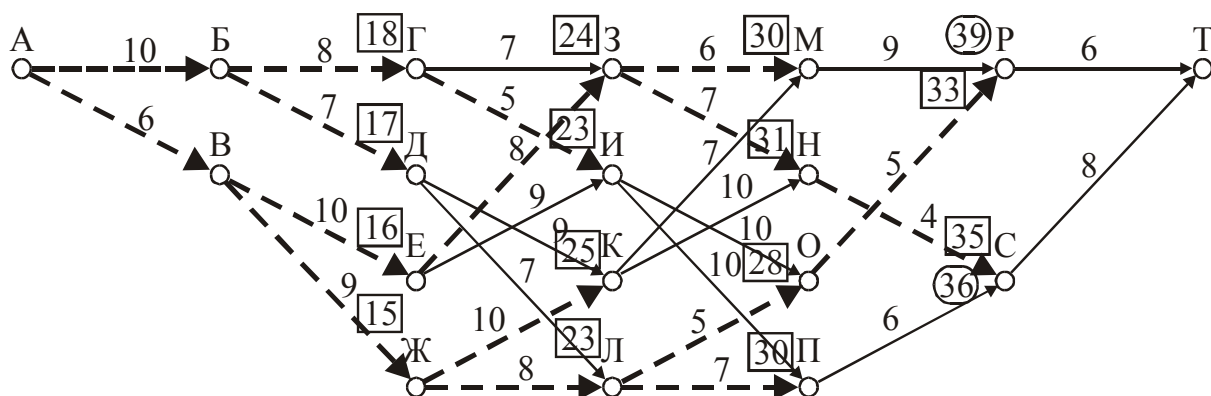


Рис. 1.7. Третий выбор альтернативных вариантов (этап 5)

Наконец, добираемся до точки назначения — города «Т», и необходимо выбрать более короткий из двух ведущих к нему путей. В данном примере эти пути отличаются на 4 км, и более коротким является верхний путь (рис. 1.8).

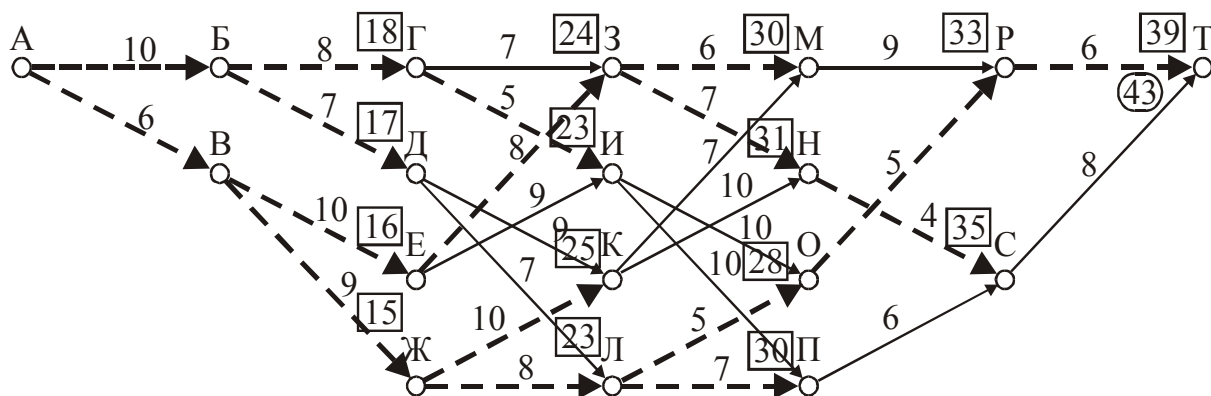


Рис. 1.8. Четвертый выбор альтернативных вариантов (этап 6)

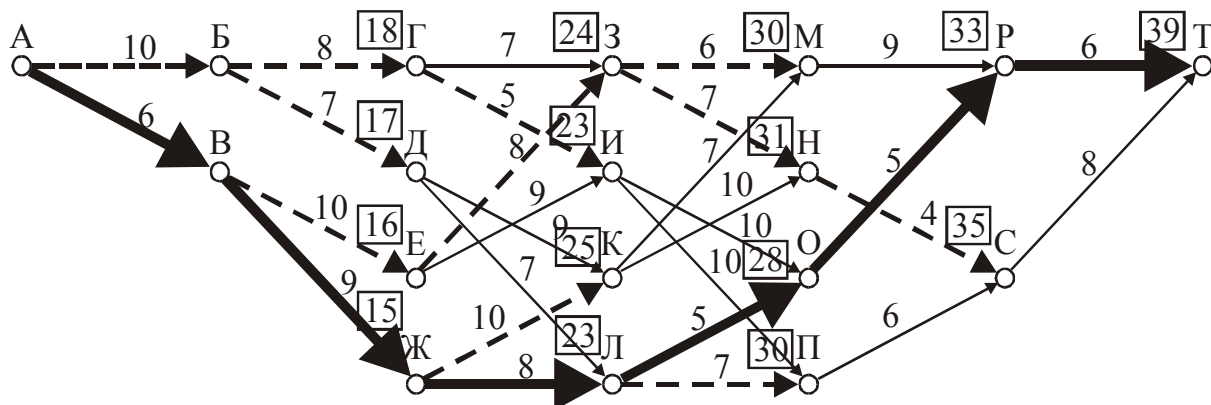


Рис. 1.9. Результат обратной трассировки

Таким образом, длина кратчайшего пути составляет 39 км, но необходимо еще определить и сам маршрут. Для этого выполняется *обратная трассировка* (backtrace). Решетка просматривается справа налево, при этом произ-

водятся переходы по выжившим путям, показанным штриховыми стрелками. Результат обратной трассировки приведен на рис. 1.9.

Итак, кратчайшее расстояние составляет 39 км, оптимальный (кратчайший) маршрут А—В—Ж—Л—О—Р—Т.

1.7. АБГШ-канал и его пропускная способность

Под термином «канал связи с аддитивным белым гауссовым шумом» (АБГШ, AWGN) подразумевается суммирование полезного сигнала $s(t)$ и белого гауссова шума $n(t)$, имеющего двустороннюю спектральную плотность $N_0/2$ (или одностороннюю СПМ N_0):

$$r(t) = s(t) + n(t).$$

Шенноном была выведена формула для пропускной способности АБГШ-канала. Пропускной способностью называется верхняя граница скоростей, при передаче данных с которыми может быть обеспечена сколь угодно низкая вероятность ошибки (при выборе соответствующих схем кодирования). В классическом виде формула Шеннона имеет следующий вид:

$$R \leq W \log_2 \left(1 + \frac{P_s}{P_n} \right) = W \log_2 (1 + \text{SNR}). \quad (1.3)$$

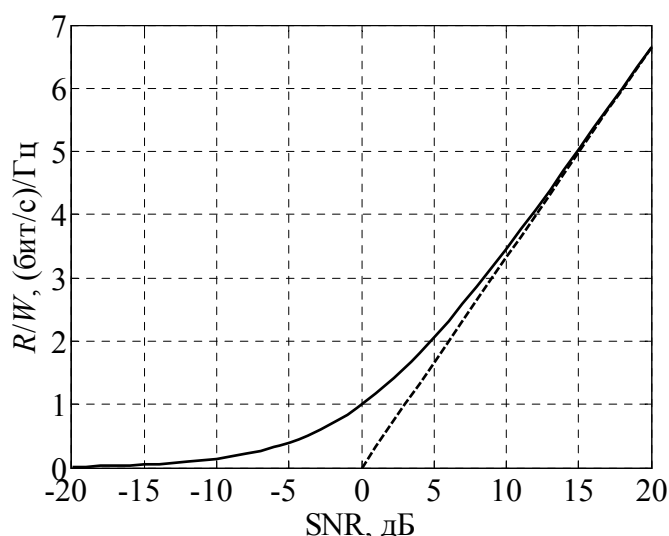


Рис. 1.10. Граница Шеннона (зависимость между спектральной эффективностью и SNR)

Таким образом, пропускная способность АБГШ-канала тем больше, чем шире его полоса частот и чем больше отношение сигнал/шум по мощности. Связь между спектральной эффективностью R/W и отношением сигнал/шум SNR иллюстрируется на рис. 1.10. Допустимые сочетания параметров SNR и R/W располагаются ниже сплошной кривой.

Из графика видно, что как спектральная эффективность, так и отношение сигнал/шум по мощности сами по себе могут принимать любые значения, однако для достижения заданной спектральной эффективности требуется определенное отношение сигнал/шум, тем большее, чем больше необходимое значение спектральной эффективности.

При больших отношениях сигнал/шум справедлива следующая асимптотика (показана на рис. 1.10 штрихами):

$$\frac{R}{W} \leq \log_2(1 + \text{SNR}) \approx \log_2 \text{SNR} = \frac{\log_2 10}{10} 10 \lg \text{SNR} \approx 0,33 \text{SNR}_{\text{дБ}},$$

где $\text{SNR}_{\text{дБ}}$ — отношение сигнал/шум по мощности в децибелах.

Однако для понимания того, что дает формула (1.3) для теории цифровой связи, полезно преобразовать ее с учетом соотношения (1.2):

$$\frac{R}{W} \leq \log_2 \left(1 + \frac{E_b}{N_0} \frac{R}{W} \right).$$

Из данной формулы легко получить неравенство, связывающее спектральную эффективность R/W и отношение сигнал/шум E_b/N_0 :

$$\frac{E_b}{N_0} \geq \frac{2^{R/W} - 1}{R/W}. \quad (1.4)$$

График, построенный согласно (1.4), показан на рис. 1.11. Допустимые сочетания параметров E_b/N_0 и R/W располагаются ниже сплошной кривой.

Сравнивая рис. 1.11 и 1.10, видим, что, в отличие от отношения сигнал/шум по мощности, отношение сигнал/шум на бит может принимать *не любые* значения. При стремлении спектральной эффективности к нулю правая часть неравенства (1.4) стремится к следующему пределу:

$$\lim_{R/W \rightarrow 0} \frac{2^{R/W} - 1}{R/W} = \ln 2 \approx 0,7 \approx -1,6 \text{ дБ}. \quad (1.5)$$

Таким образом, при $E_b/N_0 < -1,6$ дБ обеспечить безошибочную передачу в принципе невозможно, какое бы кодирование ни применялось и сколь бы низкой ни была спектральная эффективность. Данное значение называется *пределом Шеннона*.

Еще один вывод из графика на рис. 1.11 состоит в том, что работа при низких значениях E_b/N_0 возможна только при малой спектральной эффективности.

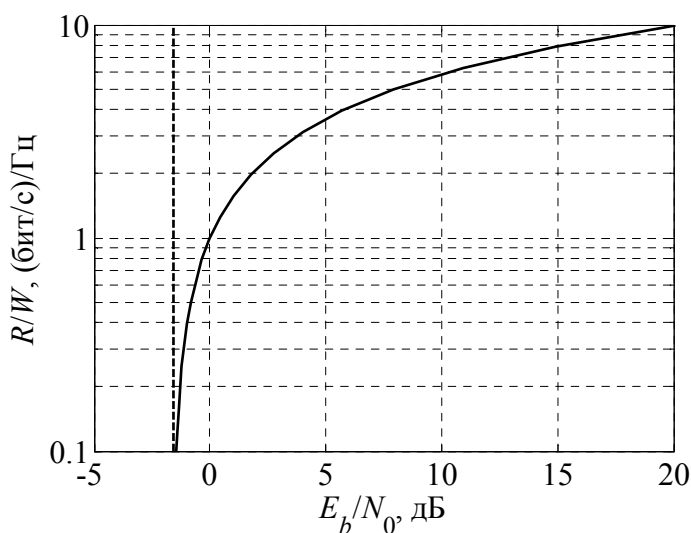


Рис. 1.11. Граница Шеннона (зависимость между спектральной эффективностью и E_b/N_0)

2. МОДУЛЯЦИЯ

Для передачи по любому каналу связи цифровое сообщение, представляющее собой последовательность символов (чисел), необходимо преобразовать в аналоговый сигнал — изменяющуюся во времени физическую величину (например, напряжение). Кроме того, канал связи способен пропускать лишь определенную полосу частот, так что сформированный аналоговый сигнал должен этой полосе соответствовать. Указанное преобразование осуществляется путем *модуляции*. Обратный процесс носит название *демодуляции*.

Для цифровых систем различают две разновидности модуляции — *низкочастотную* (baseband) и *полосовую* (passband). В первом случае формируется низкочастотный сигнал, спектр которого сосредоточен в области низких частот и занимает полосу, примерно равную символьной скорости R_{sym} (несущее колебание в данном случае не используется). Во втором случае цифровое сообщение используется для управления параметрами несущего колебания (как правило, синусоидального), таким образом, получается сигнал с полосовым спектром, сосредоточенным в окрестностях заданной несущей частоты. В данном случае символьная скорость R_{sym} также определяет ширину спектра сигнала.

Кроме того, виды цифровой модуляции делятся на 2 больших класса: линейные и нелинейные.

2.1. Линейная цифровая модуляция

2.1.1. Определение линейной модуляции

В случае линейной модуляции сам сигнал (при низкочастотной модуляции) или его комплексная огибающая (при полосовой модуляции) представляет собой *линейную комбинацию* сдвинутых по времени сигнальных посылок единой формы:

$$s(t) = \sum_{k=-\infty}^{\infty} a_k g(t - kT), \quad (2.1)$$

где T — символьный интервал; $g(t)$ — форма одиночной сигнальной посылки; a_k — коэффициенты, принимающие значения из конечного множества \mathbf{C} , называемого *сигнальным созвездием* (constellation). Эти коэффициенты и определяют передаваемую цифровую информацию, меняя лишь *масштаб* сигнальных посылок. Их часто называют *символами*, так как они являются элементами конечного множества \mathbf{C} и, следовательно, могут служить представлением передаваемых одновременно групп из нескольких бит (такое определение символа было дано в 1.4).

Так как один символ содержит m бит, возможно $M = 2^m$ различных значений символа и, следовательно, сигнальное созвездие должно содержать M точек: $\mathbf{C} = \{C_1, C_2, \dots, C_M\}$.

2.1.2. Виды созвездий, используемых на практике

В случае низкочастотной модуляции коэффициенты a_k и форма посылки $g(t)$ являются вещественными. Данный вид модуляции во многих источниках называется *амплитудно-импульсной модуляцией* (АИМ; англ. термин — Pulse Amplitude Modulation (PAM)).

Созвездие в данном случае является одномерным. При бинарной модуляции коэффициенты a_k обычно принимают значения ± 1 , при многопозиционной чаще всего используется равномерная сетка (при приведении к целочисленным значениям: $\pm 1, \pm 3, \pm 5$ и т. д.).

На рис. 2.1 показаны некоторые примеры используемых на практике форматов низкочастотной передачи.

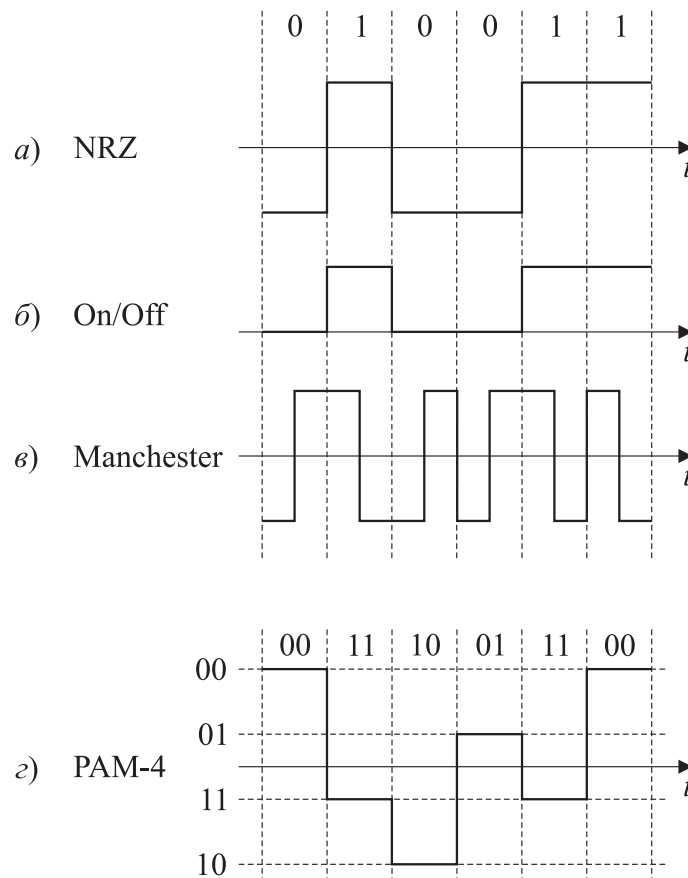


Рис. 2.1. Распространенные варианты низкочастотной модуляции

К этим распространенным форматам относятся следующие варианты:

- прямоугольные посылки, бинарная модуляция, $a_k = \pm 1$ (Non-Return to Zero, NRZ, модуляция без возврата к нулю; рис. 2.1, а): простые последовательные интерфейсы типа RS-232;
- прямоугольные посылки, однополярная бинарная модуляция, $a_k = 0$ или 1 (On/Off; рис. 2.1, б): интерфейсы для каналов связи, где невозможна инверсия полярности (например, инфракрасные каналы);
- двуполярные прямоугольные посылки с изменением полярности в середине такта, бинарная модуляция, $a_k = \pm 1$ (манчестерский код; рис. 2.1, в): сети Ethernet;
- прямоугольные посылки, многопозиционная модуляция (рис. 2.1, г): модели V.90, V.92, HDSL, SDSL.

Для случая полосовой модуляции точки созвездия C_m являются комплексными (т. е. двумерными), поэтому при выборе их размещения имеется существенно больше свободы, чем в вещественном варианте. Возможны следующие виды созвездий:

- Амплитудная модуляция (AM): $\{C_m\} = \{2m - (M - 1)\}$, $m = 0, 1, \dots, M - 1$ (на рис. 2.2 показано сигнальное созвездие AM для случая $M = 8$). На практике не используется из-за низкой помехоустойчивости, за исключением бинарных вариантов $\{C_m\} \in \{0, 1\}$ («телеграфная» модуляция, модуляция с «пассивным нулем»; англ. термин — On-Off Keying (OOK)) и $\{C_m\} \in \{-1, +1\}$ (это частный случай рассматриваемой далее фазовой модуляции).

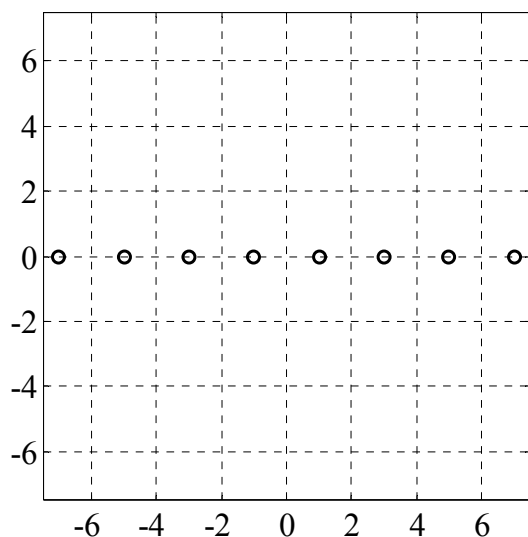


Рис. 2.2. Сигнальное созвездие 8-позиционной амплитудной модуляции

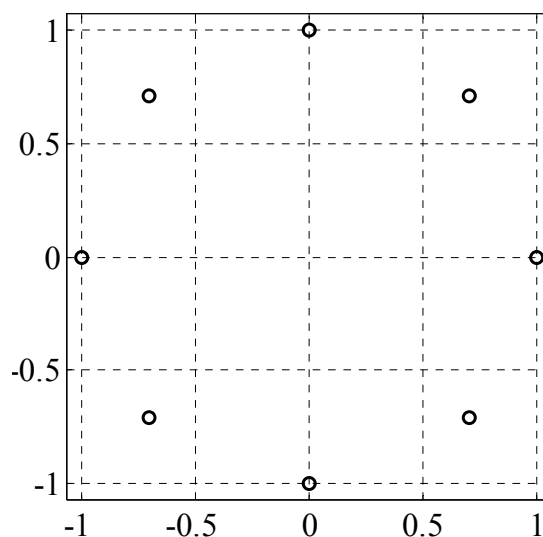


Рис. 2.3. Сигнальное созвездие 8-позиционной фазовой модуляции

- Фазовая модуляция (ФМ): $\{C_m\} = \left\{ \exp\left(j \frac{2\pi}{M} m\right) \right\}$, $m = 0, 1, \dots, M-1$ (на рис. 2.3 показано сигнальное созвездие ФМ для случая $M = 8$). На практике используется с $M = 2, 4, 8$.
- Квадратурная модуляция (КАМ): точки распределены на комплексной плоскости с использованием обеих координат (меняется как амплитуда, так и фаза). На рис. 2.4 показаны сигнальные созвездия «квадратной» КАМ для случаев $M = 16$ и $M = 32$. На практике используется с M от 4 до нескольких тысяч.

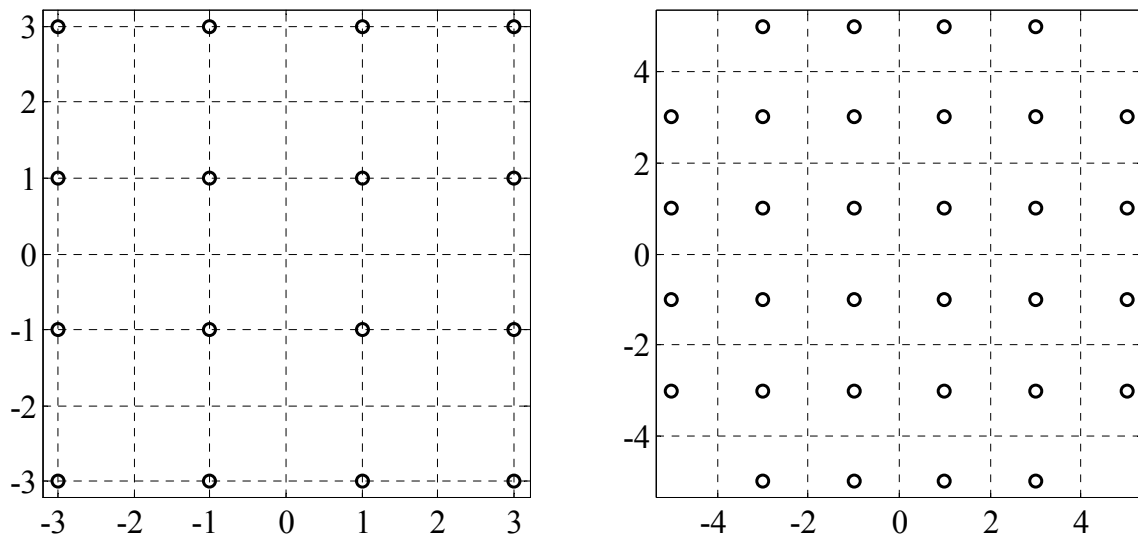


Рис. 2.4. Сигнальное созвездие 16-позиционной (слева) и 32-позиционной (справа) квадратурной модуляции

Замечание. Созвездия квадратурной модуляции не обязательно имеют «квадратный» вид, подобный показанному на рис. 2.4. «Квадратные» созвездия наиболее просто реализуются на практике, но используются и другие конфигурации. По этой причине наряду с термином «квадратурная модуляция» в литературе встречается также термин «амплитудно-фазовая модуляция» (АФМ; англ. термин — Amplitude-Phase Modulation (APM)). В некоторых источниках термин «КАМ» применяется только к созвездиям, образующим квадратную сетку, а в прочих случаях используется термин «АФМ». В данном издании термин «КАМ» будет применяться для обозначения всех созвездий линейной модуляции, при использовании которых изменяется как амплитуда, так и фаза сигнала.

2.1.3. Корреляционные и спектральные свойства сигнала с линейной модуляцией

Прежде всего необходимо отметить, что случайный сигнал, описываемый формулой (2.1), не является стационарным случайным процессом — из

за фиксированного расположения начала отсчета времени относительно сигнальных посылок он относится к категории *циклостационарных* процессов. Чтобы сделать его стационарным, необходимо ввести еще один случайный параметр — временной сдвиг t_0 , равномерно распределенный в пределах символьного интервала $0 \dots T$:

$$s(t) = \sum_{k=-\infty}^{\infty} a_k g(t - kT - t_0).$$

Вычислим корреляционную функцию (КФ) данного сигнала:

$$R(\tau) = \overline{s(t)s^*(t-\tau)} = \sum_{k_1=-\infty}^{\infty} a_{k_1} g(t - k_1T - t_0) \left(\sum_{k_2=-\infty}^{\infty} a_{k_2} g(t - k_2T - t_0 - \tau) \right)^*.$$

Прежде всего используем стандартный для подобных ситуаций прием: преобразуем произведение двух сумм в двойную сумму:

$$R(\tau) = \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} \overline{a_{k_1} a_{k_2}^*} \overline{g(t - k_1T - t_0) g^*(t - k_2T - t_0 - \tau)}.$$

Теперь произведем усреднение по передаваемым символам a_{k_1} , a_{k_2} . Поскольку предполагается, что биты в передаваемом битовом потоке являются статистически независимыми, то статистически независимыми будут и собранные из этих бит символы. Поэтому

$$\overline{a_{k_1} a_{k_2}^*} = \begin{cases} \overline{a_{k_1} a_{k_2}^*} = 0, & k_1 \neq k_2, \\ \overline{|a_{k_1}|^2} = \sigma_a^2, & k_1 = k_2. \end{cases} \quad (2.2)$$

Здесь предполагалось, что среднее значение точек созвездия равно нулю: $\overline{a_k} = \frac{1}{M} \sum_{n=1}^M A_n = 0$. Кроме того, введено обозначение σ_a^2 для среднего

квадрата модуля точек созвездия: $\sigma_a^2 = \frac{1}{M} \sum_{n=1}^M A_n^2$.

С учетом (2.2) двойная сумма превращается в одиночную:

$$R(\tau) = \sigma_a^2 \sum_{k=-\infty}^{\infty} \overline{g(t - kT - t_0) g^*(t - kT - t_0 - \tau)}.$$

Теперь необходимо выполнить усреднение по временному сдвигу t_0 (напомним, что он равномерно распределен в пределах от 0 до T):

$$R(\tau) = \sigma_a^2 \sum_{k=-\infty}^{\infty} \frac{1}{T} \int_0^T g(t - kT - t_0) g^*(t - kT - t_0 - \tau) dt_0.$$

Выполнив в интеграле замену переменной $t' = t - kT - t_0$, получаем

$$R(\tau) = \frac{\sigma_a^2}{T} \sum_{k=-\infty}^{\infty} \int_{t-kT-T}^{t-kT} g(t') g^*(t' - \tau) dt'. \quad (2.3)$$

После такого преобразования становится наглядно видно, что в (2.3) суммируются интегралы, взятые вдоль примыкающих вплотную друг к другу фрагментов временной оси длиной в символьный интервал T . Таким образом, сочетание данных интеграла и бесконечной суммы эквивалентно интегрированию вдоль всей временной оси:

$$R(\tau) = \frac{\sigma_a^2}{T} \int_{-\infty}^{\infty} g(t') g^*(t' - \tau) dt'.$$

Полученный интеграл представляет собой не что иное, как корреляционную функцию сигнальной посылки $B_g(\tau)$. Таким образом, окончательный результат можно записать следующим образом:

$$R(\tau) = \frac{\sigma_a^2}{T} B_g(\tau).$$

Далее по теореме Винера—Хинчина [4] получаем СПМ сигнала с линейной модуляцией:

$$S(f) = \frac{\sigma_a^2}{T} |\dot{G}(f)|^2,$$

где $\dot{G}(f)$ — спектральная функция сигнальной посылки:

$$\dot{G}(f) = \int_{-\infty}^{\infty} g(t) e^{-j2\pi ft} dt.$$

Таким образом, КФ и СПМ сигнала с линейной модуляцией определяются соответственно КФ и энергетическим спектром используемой сигнальной посылки.

Из полученных формул вытекают также простые соотношения для средней мощности сигнала и средней энергии одного символа:

$$\begin{aligned} P_s &= R(0) = \frac{\sigma_a^2}{T} E_g, \\ E_{\text{sym}} &= P_s T = \sigma_a^2 E_g. \end{aligned} \quad (2.4)$$

Здесь $E_g = \int_{-\infty}^{\infty} |g(t)|^2 dt$ — энергия сигнальной посылки.

2.1.4. Оптимальный прием сигналов с линейной модуляцией в АБГШ-канале

Отвлечемся на время от того факта, что символы непрерывно следуют друг за другом. Пусть передан только один символ:

$$s(t) = a_0 g(t).$$

Принятый сигнал искажен АБГШ:

$$y(t) = a_0 g(t) + n(t).$$

Задача приема сигнала с цифровой модуляцией сводится к задаче различения M сигналов. Оптимальный (по методу максимального правдоподобия (МП)) алгоритм различения в случае АБГШ-канала может быть реализован в виде коррелятора или согласованного фильтра (СФ) [1, 3.2, 4.3, прил. Б.2]. Форма опорного сигнала коррелятора (форма импульсной характеристики согласованного фильтра) согласуется с формой сигнальной посылки $g(t)$.

Замечание. Для упрощения формул импульсная характеристика СФ в дальнейшем будет рассматриваться без учета задержки, неизбежно вносимой при обработке сигнала любой физически реализуемой системой. Таким образом, будем считать, что импульсная характеристика фильтра, согласованного с сигнальной посылкой $g(t)$, равна $g(-t)$.

После получения результата вычисления корреляции (отсчета сигнала на выходе СФ) ищется точка сигнального созвездия, ближайшая к полученному значению.

2.1.5. Оценка вероятности символьной ошибки

На выходе СФ (коррелятора) получаем отсчет \dot{z} :

$$\begin{aligned} \dot{z} &= \int_{-\infty}^{\infty} y(t) g^*(t) dt = \int_{-\infty}^{\infty} (a g(t) + n(t)) g^*(t) dt = \\ &= a \int_{-\infty}^{\infty} |g(t)|^2 dt + \int_{-\infty}^{\infty} n(t) g^*(t) dt = a E_g + \int_{-\infty}^{\infty} n(t) g^*(t) dt. \end{aligned} \quad (2.5)$$

Первое слагаемое в (2.5) представляет собой вклад от полезного сигнала, второе — шум. Легко заметить, что шумовая составляющая имеет нулевое среднее значение:

$$\overline{\int_{-\infty}^{\infty} n(t) g^*(t) dt} = \int_{-\infty}^{\infty} \overline{n(t)} g^*(t) dt = 0.$$

Таким образом, математическое ожидание величины \dot{z} определяется первым слагаемым:

$$\overline{\dot{z}} = a E_g, \quad (2.6)$$

а ее дисперсия равна среднему квадрату модуля второго (шумового) слагаемого. Для вычисления преобразуем квадрат модуля от интеграла в двойной интеграл:

$$\sigma_z^2 = \overline{\left| \int_{-\infty}^{\infty} n(t)g^*(t)dt \right|^2} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \overline{n(t_1)n^*(t_2)}g^*(t_1)g(t_2)dt_1dt_2.$$

Поскольку рассматривается белый шум $n(t)$, его КФ, получаемая в результате усреднения под знаком интеграла в последней формуле, равна

$$\overline{n(t_1)n^*(t_2)} = N_0\delta(t_1 - t_2).$$

С учетом этого, используя фильтрующее свойство дельта-функции, получаем

$$\sigma_z^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} N_0\delta(t_1 - t_2)g^*(t_1)g(t_2)dt_1dt_2 = N_0 \int_{-\infty}^{\infty} |g(t)|^2 dt = N_0E_g. \quad (2.7)$$

Замечание. Здесь предполагается, что шум является *комплексным* случайным процессом. Полученное значение является *полной* дисперсией комплексной случайной величины. Вещественная и мнимая составляющая данной случайной величины имеют одинаковые дисперсии, равные $N_0E_g/2$. Такой же результат ($N_0E_g/2$) будет получен и в случае, если рассматривать *вещественный* шум. По той же причине в используемых в данном издании формулах величина N_0 обозначает *одностороннюю* СПМ в случае рассмотрения *вещественного* шума и *двустороннюю* СПМ, если рассматриваемый шум является *комплексным*.

Так как преобразования, производимые над сигналом, являются линейными, итоговая величина \dot{z} сохранит нормальное распределение. Кроме того, из (2.6) и (2.7) видно, что статистические свойства величины \dot{z} , а значит, и помехоустойчивость приема не зависят от *формы* сигнальной посылки, а определяются только ее энергией E_g .

Зная статистические свойства выходного сигнала СФ (коррелятора), можно рассчитать вероятность символьной ошибки для произвольного сигнального созвездия. Обозначим последовательность действий при таком расчете:

1. Для каждой точки созвездия определяется «область принятия решения» — часть комплексной плоскости, для точек которой данная точка созвездия будет ближайшей.
2. Для каждой точки созвездия вычисляется вероятность принятия ошибочного решения. Для расчета этой вероятности необходимо проинтегрировать двумерную гауссову плотность вероятности (ПВ) с математическим

ожиданием в точке созвездия по области, лежащей за пределами «области принятия решения», определенной на предыдущем шаге.

3. Для разных точек созвездия вероятности ошибки в общем случае будут разными. Поэтому для получения *средней* вероятности символьной ошибки результаты, полученные на предыдущем шаге, необходимо усреднить по точкам созвездия.

Приведенный алгоритм дает точный результат для вероятности символьной ошибки, но полученные формулы, как правило, являются весьма громоздкими, а в ряде случаев аналитические результаты вообще не могут быть получены, так как необходимое интегрирование можно выполнить только численными методами. Поэтому для практических целей широко используются различные приближенные и асимптотические соотношения. Рассмотрим одну из таких формул. Она получается в предположении, что перепутываться могут только *ближайшие* друг к другу пары точек созвездия. При большом отношении сигнал/шум вероятность перепутывания более далеких пар точек становится пренебрежимо малой и получаемые указанным образом значения приближаются к истинной кривой помехоустойчивости.

Пусть минимальное расстояние между точками созвездия равно d_{\min} . Тогда с учетом масштабирования выходного сигнала СФ (коррелятора) (см. (2.6)), это же расстояние для выходного сигнала станет равным $d_{\min} E_g$. Дисперсия выходного комплексного шума, согласно (2.7), равна $N_0 E_g$, но в данном случае интерес представляет только та составляющая шума, которая направлена на комплексной плоскости вдоль прямой, соединяющей рассматриваемую пару точек. Дисперсия этой составляющей в 2 раза меньше: $N_0 E_g/2$.

Ошибка при принятии решения произойдет, если отклонение, вызванное шумом, превысит половину расстояния между точками, т. е. $d_{\min} E_g/2$. Шум является нормальным, поэтому вероятность перепутывания пары точек может быть выражена следующим образом:

$$P_{d_{\min}} = 1 - \Phi\left(\frac{d_{\min} E_g/2}{\sqrt{N_0 E_g/2}}\right) = 1 - \Phi\left(\sqrt{\frac{d_{\min}^2 E_g}{2N_0}}\right) = Q\left(\sqrt{\frac{d_{\min}^2 E_g}{2N_0}}\right). \quad (2.8)$$

Здесь $\Phi(x)$ — гауссов интеграл вероятности, а $Q(x)$ — его дополнение до единицы, получившее в теории связи название Q -функции:

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt, \quad Q(x) = 1 - \Phi(x) = \Phi(-x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

Далее учтем в (2.8) формулу (2.4), связывающую энергию сигнальной посылки и среднюю энергию символа:

$$P_{d_{\min}} = Q \left(\sqrt{\frac{d_{\min}^2}{2\sigma_a^2} \frac{E_{\text{sym}}}{N_0}} \right).$$

Теперь в формуле появилось *отношение сигнал/шум на символ* — E_{sym}/N_0 .

Чтобы рассчитать полную вероятность символьной ошибки, необходимо умножить $P_{d_{\min}}$ на удвоенное число пар точек с минимальным расстоянием (n_{\min}), имеющих в сигнальном созвездии, и разделить результат на размер созвездия M . Удвоение необходимо, так как каждая пара принимает участие в двух «зеркальных» событиях перепутывания. Деление на M производит усреднение по точкам созвездия. В итоге получаем

$$P_{\text{sym}} \approx \frac{2n_{\min}}{M} P_{d_{\min}} = \frac{2n_{\min}}{M} Q \left(\sqrt{\frac{d_{\min}^2}{2\sigma_a^2} \frac{E_{\text{sym}}}{N_0}} \right). \quad (2.9)$$

Чтобы перейти в (2.9) к *отношению сигнал/шум на бит*, необходимо вспомнить, что символ состоит из m бит и, следовательно, $E_{\text{sym}} = m E_b$:

$$P_{\text{sym}} \approx \frac{2n_{\min}}{M} Q \left(\sqrt{\frac{d_{\min}^2}{2\sigma_a^2} \frac{mE_b}{N_0}} \right). \quad (2.10)$$

При большом отношении сигнал/шум действительно почти всегда перепутываются только ближайшие пары точек и результат (2.10) приближается к точному значению. При снижении отношения сигнал/шум начинает заметно проявляться вероятность перепутывания все более удаленных друг от друга пар точек и истинная вероятность ошибки становится больше, чем получается по формуле (2.10). При дальнейшем снижении отношения сигнал/шум начинает сказываться тот факт, что формула (2.10), по сути дела, получена частичным вычислением *аддитивной границы* для вероятности объединения различных событий перепутывания. Аддитивная граница является верхней границей и, таким образом, дает завышенный результат. Поэтому и результат (2.10) при малых отношениях сигнал/шум может оказываться завышенным. В частности, из формулы видно, что при отношении сигнал/шум, равном нулю, получаемый результат, равный n_{\min}/M , может превышать не только $(M-1)/M$, но и единицу.

При использовании бинарной фазовой модуляции (ФМ-2, BPSK) созвездие содержит всего две точки (+1 и -1), поэтому формула (2.10) в данном случае является точной. Для ее применения к этому виду модуляции необходимо подставить следующие значения: $m = 1$, $M = 2$, $d_{\min} = 2$, $\sigma_a^2 = 1$, $n_{\min} = 1$.

Так как модуляция является бинарной, полученная формула дает вероятность как символьной, так и битовой ошибки:

$$P_b = P_{\text{sym}} = Q\left(\sqrt{\frac{2E_b}{N_0}}\right). \quad (2.11)$$

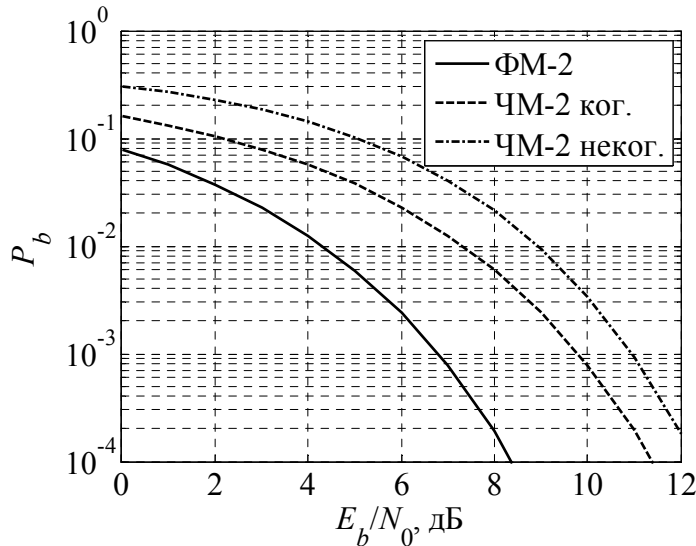


Рис. 2.5. Кривые помехоустойчивости для модуляции ФМ-2 (сплошная линия), ортогональной ЧМ-2 с когерентным (штрихи) и некогерентным (штрихпунктир) приемом

Кривая помехоустойчивости для модуляции ФМ-2 показана на рис. 2.5 сплошной линией. Две остальные кривые соответствуют приему частотной модуляции (см. 2.2.4 и 2.2.5).

Из формулы (2.10) видно, что асимптотическая помехоустойчивость линейной модуляции определяется двумя коэффициентами — множителем d_{\min}/σ_a в аргументе Q -функции и множителем n_{\min}/M при самой Q -функции. Однако, так как затухание Q -функции с ростом аргумента носит экспоненциальный характер, влияние

множителя d_{\min}/σ_a сказывается на помехоустойчивости значительно сильнее.

Таким образом, из двух созвездий с одинаковым числом точек лучшую помехоустойчивость обеспечит то, у которого больше отношение d_{\min}/σ_a .

2.1.6. Оценка вероятности битовой ошибки. Код Грея

Вероятность символьной ошибки определяется только отношением сигнал/шум и геометрией сигнального созвездия. Вероятность же *битовой* ошибки, кроме этих факторов, зависит еще и от того, каким именно образом m -битовые комбинации размещены по точкам созвездия. В общем случае, для произвольного созвездия и произвольного назначения этих m -битовых комбинаций, процедура *точного* расчета вероятности битовой ошибки должна выглядеть следующим образом:

1. Для каждой из M точек сигнального созвездия по отдельности определяется вероятность принятия решения в пользу *каждой* из оставшихся $(M - 1)$ точек и фиксируется число ошибочных бит, возникающих при таком решении.
2. Полученные на предыдущем шаге $M(M - 1)$ значений числа ошибочных бит усредняются с учетом вероятностей соответствующих событий. Ре-

зультат дает среднюю вероятность битовой ошибки для данного отношения сигнал/шум.

При большом отношении сигнал/шум, когда преимущественно перепутываются ближайшие точки созвездия, получить приближенное значение вероятности битовой ошибки можно следующим образом. Вычисляется среднее значение числа бит, отличающихся в парах ближайших друг к другу точек созвездия ($b_{\text{ср}}$), делится на общее число бит в символе (m) и умножается на вероятность символьной ошибки, полученную по формуле (2.10):

$$P_b \approx \frac{b_{\text{ср}}}{m} P_{\text{sym}}. \quad (2.12)$$

Из изложенного ясна важность оптимизации размещения битовых комбинаций по точкам созвездия. Общую стратегию такой оптимизации можно качественно сформулировать следующим образом: чем ближе друг к другу расположены точки созвездия (и чем больше, следовательно, вероятность их перепутывания), тем меньшим числом бит должны отличаться присвоенные им битовые комбинации.

Большое практическое распространение получил код Грея, при использовании которого битовые комбинации для соседних точек созвездия различаются лишь одним битом. Исходя из (2.12), при большом отношении сигнал/шум и использовании кода Грея вероятности символьной и битовой ошибок приближенно связаны друг с другом следующим образом:

$$P_b \approx P_{\text{sym}}/m.$$

Учтя здесь выражение (2.10) для P_{sym} , получаем приближенное выражение для вероятности битовой ошибки, справедливое при большом отношении сигнал/шум:

$$P_b \approx \frac{2n_{\text{min}}}{m2^m} Q \left(\sqrt{\frac{d_{\text{min}}^2}{2\sigma_a^2} \frac{mE_b}{N_0}} \right).$$

Одномерный код Грея может быть построен, например, следующим итерационным способом. Представим битовые комбинации, соответствующие отдельным точкам созвездия, в виде столбцов матрицы \mathbf{B}_m размера $m \times M$ (m строк, $M = 2^m$ столбцов). Для однобитового случая раскладка очевидна:

$$\mathbf{B}_1 = [0 \mid 1].$$

Если имеется набор m -битовых комбинаций \mathbf{B}_m , расширить его до $m + 1$ бит можно следующим образом:

$$\mathbf{B}_{m+1} = \left[\begin{array}{c|c} \mathbf{B}_m & \overline{\mathbf{B}_m} \\ \hline 0 \dots 0 & 1 \dots 1 \end{array} \right].$$

Здесь $\overline{\mathbf{B}}_m$ обозначает матрицу \mathbf{B}_m , зеркально перевернутую вдоль горизонтальной координаты (вдоль строк). Доказать, что набор столбцов матрицы \mathbf{B}_{m+1} будет представлять собой код Грея, можно методом математической индукции. Кроме того, легко заметить, что крайние правый и левый столбцы матрицы также отличаются друг от друга лишь одним элементом, т. е. полученный код может быть замкнут в кольцо (это дает возможность использовать его для созвездий фазовой модуляции).

Покажем несколько первых итераций, необходимых для формирования многобитового кода Грея:

$$\mathbf{B}_2 = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad \mathbf{B}_3 = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix},$$

$$\mathbf{B}_4 = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

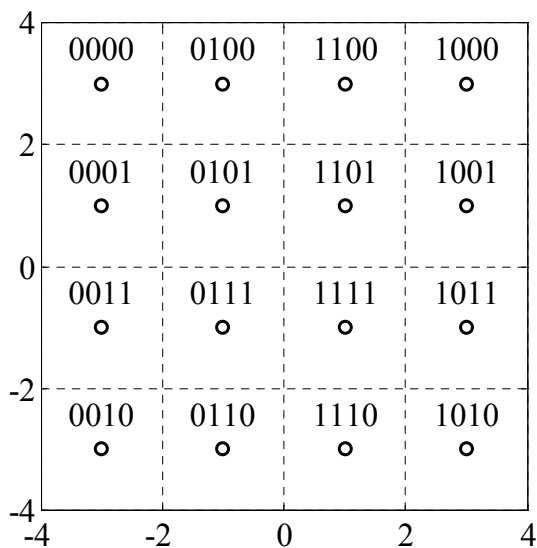


Рис. 2.6. Код Грея для созвездия КАМ-16

Полученные указанным образом наборы битовых комбинаций могут быть непосредственно использованы для созвездий амплитудной и фазовой манипуляции. В случае «квадратной» КАМ, когда символ содержит четное число бит, вдоль каждой из координат реализуется $m/2$ -битовый код Грея. Легко заметить, что ближайшие точки двумерного КАМ-созвездия в этом случае будут также отличаться лишь одним битом. Пример битовой раскладки в соответствии с кодом Грея для созвездия КАМ-16 приведен на рис. 2.6.

Что касается «крестообразных» КАМ-созвездий с нечетным числом бит в символе, для них также возможно построение двумерного кода Грея, но более сложным образом. Кроме того, в ряде случаев при этом не удастся сформировать код так, чтобы соседние точки отличались только одним битом; часть таких пар будет отличаться двумя битами.

2.1.7. Формирование спектра и условие отсутствия межсимвольной интерференции

В 2.1.5 было показано, что помехоустойчивость линейной модуляции при оптимальном приеме не зависит от формы сигнальной посылки, но в то же время эта форма определяет корреляционные и спектральные свойства сигнала. Теперь вспомним тот факт, о котором на время забыли в начале подраздела 2.1.4 — символы сообщения непрерывно следуют друг за другом. Посмотрим, какие дополнительные ограничения это накладывает на возможные формы сигнальной посылки.

После пропускания сигнала (2.1) через СФ он принимает следующий вид (для упрощения выкладок здесь не учтена задержка фильтрации):

$$y(t) = \sum_{k=-\infty}^{\infty} a_k B_g(t - kT), \quad (2.13)$$

где $B_g(t)$ — корреляционная функция импульса $g(t)$. Согласно теории оптимального приема, необходимо брать отсчеты этого сигнала в тех точках, где корреляционная функция имеет максимум, т. е. в моменты времени $t_n = nT$:

$$y(t_n) = \sum_{k=-\infty}^{\infty} a_k B_g((n - k)T). \quad (2.14)$$

Главной особенностью систем связи является то, что сигнальные посылки следуют друг за другом с интервалом T и при приеме не должны мешать друг другу. Эти взаимные помехи называются межсимвольной интерференцией (МСИ; англ. термин — InterSymbol Interference (ISI)). Из (2.14) следует условие отсутствия МСИ:

$$B_g(kT) = 0, \quad k = \pm 1, \pm 2, \dots \quad (2.15)$$

Отсюда сразу же следует, что МСИ будет отсутствовать при любой форме посылки конечной длительности, если эта длительность не превышает T . Простейшей формой посылки такого рода является прямоугольный импульс:

$$g(t) = \begin{cases} 1, & 0 \leq t \leq T, \\ 0, & t < 0, \quad t > T. \end{cases} \quad (2.16)$$

Как было показано в 2.1.3, форма используемого импульса определяет спектральные свойства сформированного сигнала, так что СПМ получившегося сигнала пропорциональна энергетическому спектру формы одиночной посылки. Прямоугольные импульсы — не самый лучший выбор, так как спектр прямоугольного импульса затухает медленно и у него высок уровень боковых лепестков (уровень первого лепестка -13 дБ). По этой причине на практике используются посылки непрямоугольной формы, обладающие более компактным спектром.

Из условия (2.15), сформулированного во временной области, неочевидно, что МСИ может отсутствовать и при посылках, не укладывающихся в символьный интервал. Чтобы показать это, необходимо сформулировать условие отсутствия МСИ в частотной области.

Согласно (2.15), КФ посылки, дискретизированная по времени с шагом T , представляет собой одиночный дискретный импульс:

$$B_g(kT) = \begin{cases} E_g, & k = 0, \\ 0, & k \neq 0. \end{cases} \quad (2.17)$$

Как известно из теории дискретных сигналов (см., например, [4, 6.5]), спектр дискретизированного сигнала пропорционален бесконечной сумме сдвинутых на частоту дискретизации копий спектра исходного аналогового сигнала. С другой стороны, спектр одиночного дискретного импульса (2.17) представляет собой константу. С учетом этого получаем условие отсутствия МСИ, записанное в частотной области:

$$\sum_{n=-\infty}^{\infty} \left| \dot{G}\left(f - \frac{n}{T}\right) \right|^2 = \text{const}. \quad (2.18)$$

Условие (2.18) позволяет находить подходящие формы посылок, имеющих бесконечную длительность. В частности, из него очевидно, что МСИ будет отсутствовать, если $g(t)$ обладает прямоугольным низкочастотным спектром, занимающим полосу частот от нуля до $\pm 1/(2T)$. Сама функция $g(t)$ при этом имеет вид $\sin(x)/x$, представляя собой базисную функцию ряда Котельникова (рис. 2.7):

$$g(t) = \sqrt{\frac{E_g}{T}} \frac{\sin(\pi t/T)}{\pi t/T}. \quad (2.19)$$

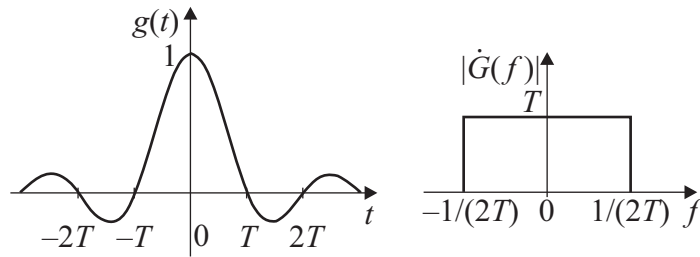


Рис. 2.7. Сигнальная посылка с минимально возможной шириной спектра

Из рис. 2.7, в частности, видна минимально достижимая ширина спектра сигнала с линейной частотной модуляцией — при низкочастотной модуляции она равна половине символьной скорости ($1/(2T) = R_{\text{sym}}/2$), а при полосовой оказывается вдвое больше: $1/T = R_{\text{sym}}$. Указанные величины чаще всего и подразумеваются под шириной спектра сигнала с линейной модуляцией.

В частности, при таком определении ширины спектра получается, что спектральная эффективность полосовой линейной модуляции равна $m = \log_2 M$, т. е. числу бит, содержащихся в одном символе.

Импульс (2.19) дает минимальную ширину спектра сигнала, но он является бесконечным по длительности и его «хвосты» затухают медленно. Поэтому наибольшее практическое распространение получил другой импульс, обладающий большей скоростью спада во временной области. Соответствующая форма спектра называется *косинусоидально-сглаженной* (raised cosine), при этом край спектра $|\dot{G}(f)|^2$ «разглаживается» по косинусоидальному закону (рис. 2.8):

$$|\dot{G}(f)|^2 = E_g T \times \begin{cases} 1, & |2fT| \leq 1 - \alpha, \\ \frac{1}{2} - \frac{1}{2} \sin\left(\frac{\pi}{2} \frac{|2fT| - 1}{\alpha}\right), & 1 - \alpha < |2fT| \leq 1 + \alpha, \\ 0, & |2fT| > 1 + \alpha. \end{cases} \quad (2.20)$$

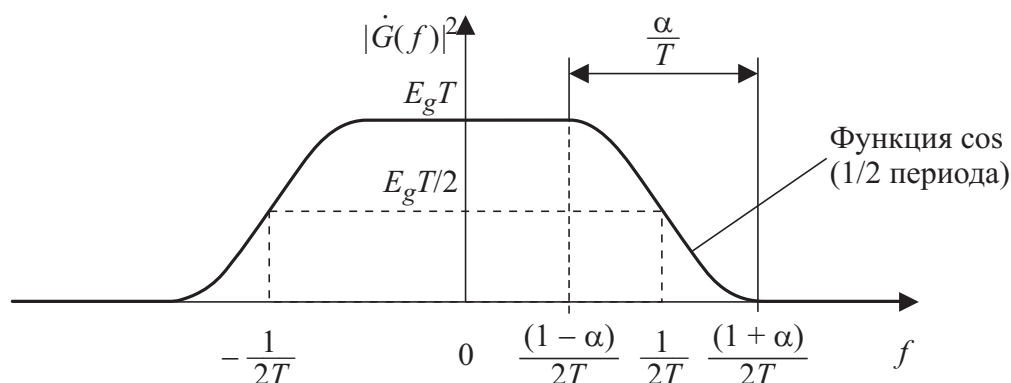


Рис. 2.8. АЧХ с косинусоидальным сглаживанием

Параметр α называется *коэффициентом сглаживания* (rolloff factor), он равен половине ширины переходной зоны, нормированной к частоте среза. При $\alpha = 0$ фильтр превращается в идеальный ФНЧ с прямоугольной АЧХ (см. (2.19) и рис. 2.7), при $\alpha = 1$ в АЧХ отсутствует плоский участок в полосе пропускания.

Обратное преобразование Фурье дает выражение для КФ рассматриваемой сигнальной посылки:

$$B_g(\tau) = E_g \frac{\cos(\alpha\pi\tau/T)}{1 - (2\alpha\tau/T)^2} \frac{\sin(\pi\tau/T)}{\pi\tau/T}.$$

Использованный способ сглаживания делает непрерывной не только саму АЧХ, но и ее первую производную, поэтому полученная КФ убывает пропорционально τ^3 .

На рис. 2.9 показаны графики $B_g(\tau)$ при разных значениях коэффициента сглаживания α . Видно, что с ростом α боковые лепестки функции спадают быстрее (платой за это является расширение спектра сигнала).

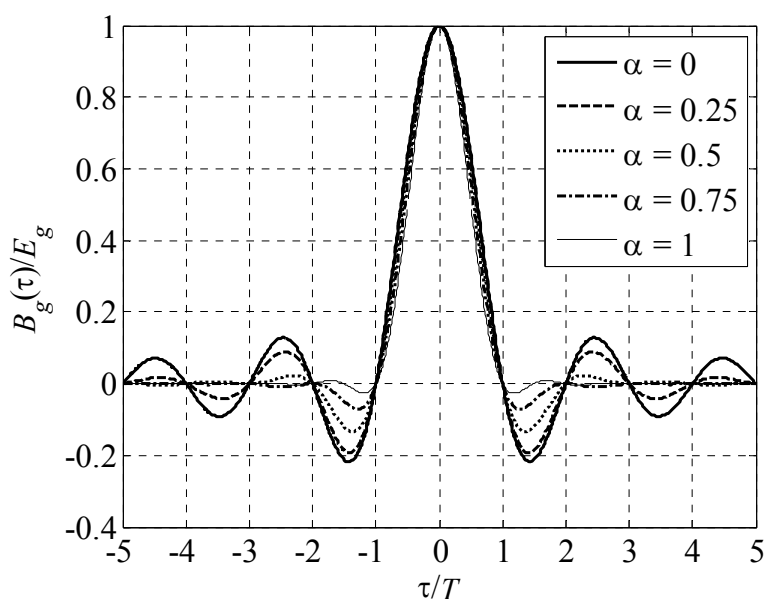


Рис. 2.9. КФ импульсной характеристики фильтра с косинусоидальным сглаживанием АЧХ

АЧХ формирующего фильтра в передатчике и СФ в приемнике (рис. 2.10) будет представлять собой квадратный корень из формулы (2.20):

$$\dot{G}(f) = \sqrt{E_g T} \times \begin{cases} 1, & |2fT| \leq 1 - \alpha, \\ \cos\left(\frac{\pi}{4} \frac{|2fT| - (1 - \alpha)}{\alpha}\right), & 1 - \alpha < |2fT| \leq 1 + \alpha, \\ 0, & |2fT| > 1 + \alpha. \end{cases}$$

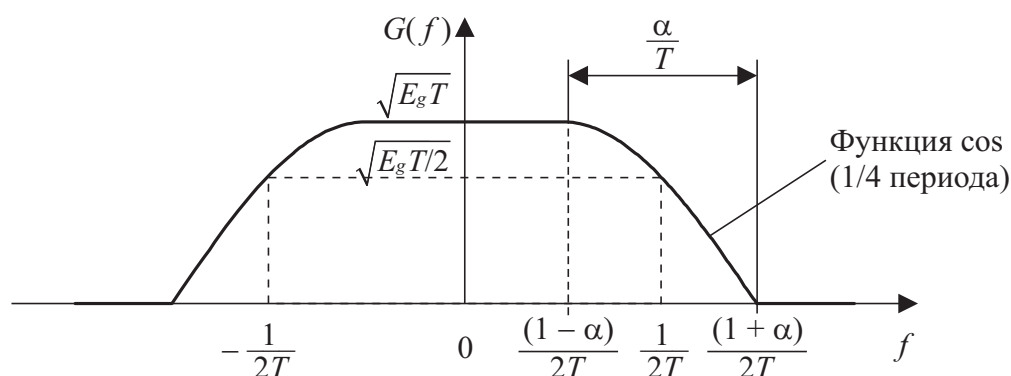


Рис. 2.10. АЧХ SQRT-варианта ФНЧ с косинусоидальным сглаживанием

Импульсная характеристика такого фильтра описывается следующей формулой (фазочастотная характеристика подразумевается равной нулю):

$$g(t) = \sqrt{\frac{E_g}{T}} \frac{4\pi\alpha}{\pi^2 - (4\pi\alpha t/T)^2} \left(\cos((1+\alpha)\pi t/T) + \frac{\sin((1-\alpha)\pi t/T)}{4\alpha t/T} \right).$$

Фильтр такого типа называется *SQRT-вариантом* фильтра с косинусоидальным сглаживанием АЧХ (square root raised cosine filter). Его импульсная характеристика при разных значениях коэффициента сглаживания α показана на рис. 2.11.

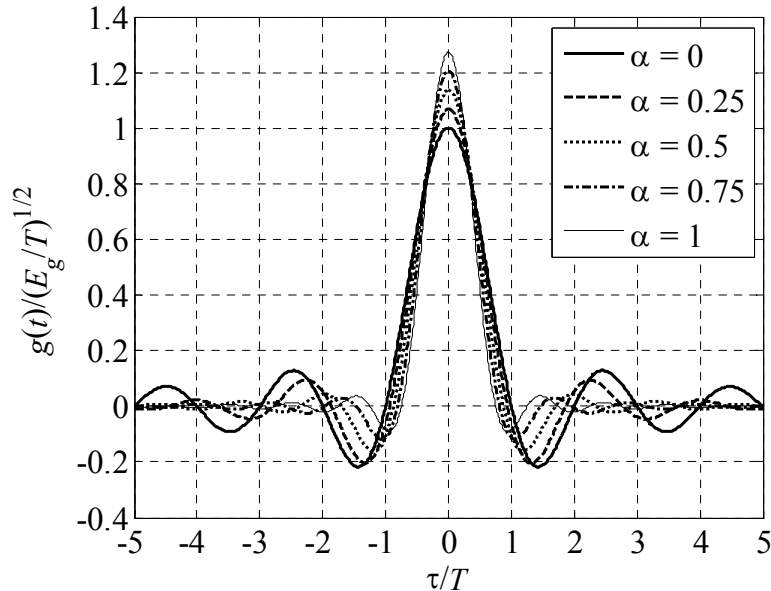


Рис. 2.11. Импульсная характеристика SQRT-фильтра с косинусоидальным сглаживанием АЧХ

Из графиков на рис. 2.11 виден важный факт: при $\alpha \neq 0$ сформированный сигнал *в любой момент времени* представляет собой сумму вкладов от нескольких символов сообщения, т. е. в нем *нельзя* выделить временные сечения, свободные от межсимвольной интерференции. Такие моменты времени появляются *только после согласованной фильтрации* сигнала.

2.2. Нелинейная цифровая модуляция

2.2.1. Определение нелинейной модуляции

Для нелинейной цифровой модуляции сложно дать столь однозначное определение, как для модуляции линейной. В частности, сложность состоит в том, что некоторые виды нелинейной модуляции могут быть представлены как линейная модуляция в сочетании с некоторым предварительным преобразованием (кодированием) символов сообщения. Такой сигнал описывается той же формулой (2.1), но из-за предварительного кодирования могут измениться статистические свойства символов a_k . Однако, все же формулируя отличие линейной цифровой модуляции от нелинейной, можно отметить, что в случае линейной модуляции передаваемые символы меняют только *масштаб*

сигнальных посылок, а при модуляции нелинейной от них зависит сама *форма* сигнальной посылки.

При низкочастотной модуляции к нелинейным видам, в частности, относятся *временн*ая импульсная и широтно-импульсная модуляция.

При *временн*ой импульсной модуляции (ВИМ; англ. термин — Pulse Position Modulation (PPM)) в каждом символьном интервале передается один импульс, положение которого в пределах символьного интервала может принимать M различных значений; это положение и определяет передаваемую информацию (рис. 2.12, а).

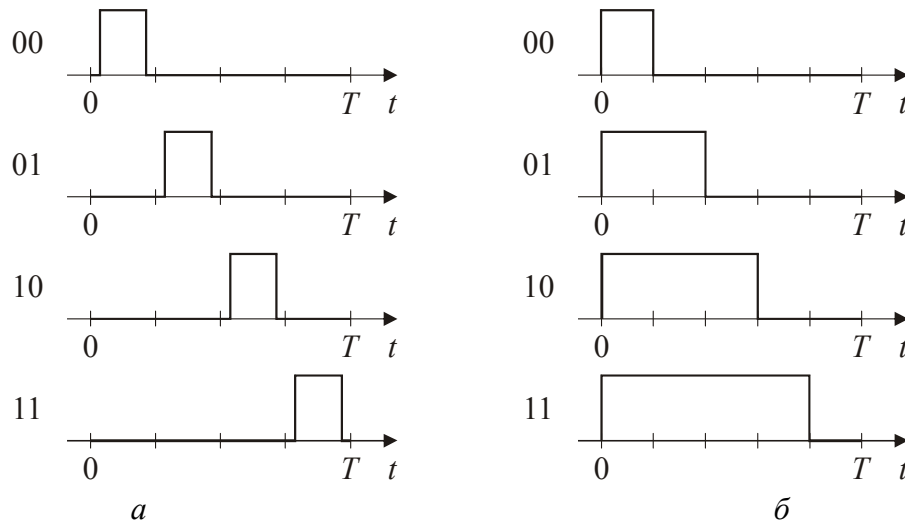


Рис. 2.12. Четырехпозиционная временн

ая импульсная (а) и широтно-импульсная (б) модуляция

При *широтно-импульсной модуляции* (ШИМ; англ. термин — Pulse Width Modulation (PWM)) в каждом символьном интервале передается один импульс, длина которого может принимать M различных значений; эта длина и определяет передаваемую информацию (см. рис. 2.12, б).

Среди видов полосовой модуляции к нелинейным относится прежде всего модуляция (или манипуляция) частотная.

2.2.2. Цифровая частотная модуляция

Сигналы с цифровой частотной модуляцией (ЧМ; англ. термин — Frequency Shift Keying (FSK)) относятся к классу сигналов с угловой модуляцией, так что комплексная огибающая такого сигнала может быть записана следующим образом:

$$\dot{s}(t) = \sqrt{\frac{E_{\text{sym}}}{T}} e^{j\varphi(t)}, \quad (2.21)$$

где E_{sym} — энергия символа; T — длительность символьного интервала; $\varphi(t)$ — фазовая функция.

В простейшем (*несглаженном*) варианте фазовая функция на каждом символьном интервале зависит от времени по линейному закону:

$$\varphi(t) = 2\pi f_k(t - kT) + \varphi_k, \quad kT \leq t < (k+1)T, \quad (2.22)$$

где f_k и φ_k — соответственно частота и начальная фаза сигнала на k -м символьном интервале. Частоты f_k несут информацию, они выбираются из некоторого набора из M частот манипуляции: $f_k \in \mathbf{F} = \{F_1, F_2, \dots, F_M\}$.

Напомним, что рассматривается комплексная огибающая сигнала, поэтому в приводимых формулах под частотами манипуляции подразумеваются частотные сдвиги относительно несущей частоты сигнала.

Объединяя отдельные символы с фазовыми функциями вида (2.22), сигнал с угловой модуляцией вида (2.21) можно записать следующим образом:

$$\dot{s}(t) = \sqrt{\frac{E_{\text{sym}}}{T}} \sum_{k=-\infty}^{\infty} g(t - kT) e^{j(2\pi f_k(t - kT) + \varphi_k)},$$

где $g(t)$ — прямоугольная посылка вида (2.16).

Чаще всего частоты манипуляции образуют равномерную сетку с некоторым шагом Δf :

$$f_n = \Delta f \left(n - \frac{M+1}{2} \right), \quad n = 1, 2, \dots, M.$$

При бинарной передаче $M = 2$, так что $f_{1,2} = \pm \Delta f / 2$.

2.2.3. Корреляционные и спектральные свойства сигналов с частотной манипуляцией

Начальные фазы отдельных символов φ_k (см. (2.22)) могут формироваться различным образом в зависимости от способа генерирования сигнала. Закон их формирования весьма существенно влияет на корреляционные и спектральные свойства получающегося ЧМ-сигнала. Рассмотрим несколько возможных вариантов.

1. Независимые начальные фазы. В этом случае начальные фазы отдельных символов φ_k представляют собой статистически независимые случайные величины с равномерным распределением на интервале $0 \dots 2\pi$. Этот вариант редко применяется (если вообще применяется) на практике, но он позволяет легко вычислить КФ и СПМ комплексной огибающей сигнала.

Рассчитаем для данного случая КФ ЧМ-сигнала (так же, как и в 2.1.3, чтобы сделать сигнал стационарным случайным процессом, необходимо добавить в математическую модель случайный временной сдвиг t_0 , равномерно распределенный на интервале $0 \dots T$). При $|\tau| > T$ КФ, очевидно, будет равна нулю, так как 2 момента времени попадают в разные символьные интервалы. При $|\tau| \leq T$ два момента времени попадают в разные символьные интервалы,

давая при этом нулевую КФ, с вероятностью $|\tau|/T$. С вероятностью $(1 - |\tau|/T)$ эти моменты попадают в один и тот же символьный интервал и дают значение КФ, определяемое следующим образом:

$$\sqrt{\frac{E_{\text{sym}}}{T}} e^{j(2\pi f_k(t-kT)+\varphi_k)} \sqrt{\frac{E_{\text{sym}}}{T}} e^{-j(2\pi f_k(t-\tau-kT)+\varphi_k)} = \frac{E_{\text{sym}}}{T} e^{j2\pi f_k \tau}.$$

Частота f_k равновероятно принимает значения из множества частот манипуляции \mathbf{F} , поэтому данный результат необходимо усреднить по частотам этого множества. Объединяя все изложенное, окончательно получаем:

$$\dot{R}(\tau) = \begin{cases} \frac{E_{\text{sym}}}{T} \left(1 - \frac{|\tau|}{T}\right) \frac{1}{M} \sum_{n=1}^M \exp(j2\pi F_n \tau), & |\tau| \leq T, \\ 0, & |\tau| > T. \end{cases} \quad (2.23)$$

Преобразование Фурье от КФ (2.23), согласно теореме Винера—Хинчина, дает СПМ комплексной огибающей. Эта СПМ представляет собой сумму сдвинутых по частоте функций вида $(\sin x / x)^2$:

$$S(f) = \frac{E_{\text{sym}}}{M} \sum_{n=1}^M \left(\frac{\sin(\pi(f - F_n)T)}{\pi(f - F_n)T} \right)^2. \quad (2.24)$$

Из (2.24) видно, что, благодаря случайным начальным фазам посылок, спектр не содержит дискретных составляющих.

2. Коммутация независимых генераторов. Если ЧМ-сигнал формируется коммутацией сигналов независимых генераторов с частотами F_n (так часто происходило в системах, построенных на базе аналоговой обработки сигналов), то фрагменты сигнала, соответствующие одинаковым значениям символа ($f_k = F_n$), когерентны по отношению друг к другу, так как они «вырезаны» из непрерывной синусоиды. Из-за этого в корреляционной функции появятся периодические составляющие, а в спектре, соответственно, дискретные компоненты (слагаемые в виде дельта-функций). Рассуждения, приведенные ранее для случайных начальных фаз, приводят к следующим результатам для КФ и СПМ (начальные фазы разных генераторов считаем случайными, равномерно распределенными на интервале $0 \dots 2\pi$ и статистически независимыми друг от друга):

$$\dot{R}(\tau) = \begin{cases} \frac{E_{\text{sym}}}{T} \left(1 - \frac{M-1}{M} \frac{|\tau|}{T}\right) \frac{1}{M} \sum_{n=1}^M \exp(j2\pi F_n \tau), & |\tau| \leq T, \\ \frac{E_{\text{sym}}}{T} \frac{1}{M^2} \sum_{n=1}^M \exp(j2\pi F_n \tau), & |\tau| > T. \end{cases}$$

$$S(f) = E_{\text{sym}} \frac{M-1}{M^2} \sum_{n=1}^M \left(\frac{\sin(\pi(f - F_n)T)}{\pi(f - F_n)T} \right)^2 + \frac{E_{\text{sym}}}{T} \frac{1}{M^2} \sum_{n=1}^M \delta(f - F_n).$$

3. ЧМ с непрерывной фазой. В обоих рассмотренных ранее вариантах при переходе от одного символьного такта к другому фазовая функция может изменяться скачкообразно. Поэтому сам ЧМ-сигнал тоже может испытывать разрывы, и, в соответствии со свойствами преобразования Фурье, его СПМ затухает медленно — пропорционально лишь второй степени частоты. Следующий способ формирования начальных фаз лишен этого недостатка.

Мгновенная частота сигнала является, как известно, производной от его фазовой функции. Это дает возможность формировать ЧМ-сигнал следующим образом: сначала формируется сигнал, задающий зависимость мгновенной частоты от времени, затем этот сигнал интегрируется и получается фазовая функция $\varphi(t)$, которая, в свою очередь, используется для формирования самого ЧМ-сигнала (2.21). При таком способе формирования фазовая функция оказывается непрерывной, следовательно, является непрерывным и сам сигнал, а это значит, что его СПМ будет затухать существенно быстрее — пропорционально уже как минимум четвертой степени частоты. Такая разновидность цифровой частотной модуляции называется *ЧМ с непрерывной фазой* (ЧМНФ; англ. термины — Continuous Phase Modulation (CPM) и Continuous Phase Frequency Shift Keying (CPFSK)).

Математически ЧМНФ-сигнал описывается следующим образом. *Мгновенная частота* сигнала представляет собой сигнал с амплитудно-импульсной модуляцией (см. рис. 2.1):

$$f(t) = \sum_{k=-\infty}^{\infty} a_k q(t - kT),$$

где a_k — значения, соответствующие символам передаваемого сообщения (как правило, используются целочисленные значения $\pm 1, \pm 3, \dots, \pm(M-1)$); $q(t)$ — форма *частотной посылки*.

Интегрирование данного сигнала по времени дает фазовую функцию (с точностью до постоянной начальной фазы):

$$\varphi(t) = 2\pi \int_{-\infty}^t f(t') dt' = 2\pi \int_{-\infty}^t \left(\sum_{k=-\infty}^{\infty} a_k q(t' - kT) \right) dt' = \sum_{k=-\infty}^{\infty} a_k \phi(t - kT), \quad (2.25)$$

где $\phi(t) = 2\pi \int_{-\infty}^t q(t') dt'$ — форма *фазового перехода*, вызываемого одним передаваемым символом.

В рассматриваемом до сих пор случае несглаженной ЧМ частотная посылка $q(t)$ имеет прямоугольную форму, а фазовый переход $\phi(t)$ является кусочно-линейной функцией:

$$q(t) = \begin{cases} \Delta f / 2, & 0 \leq t < T, \\ 0 & t < 0, \quad t \geq T, \end{cases} \quad \phi(t) = \begin{cases} 0, & t < 0, \\ \pi \Delta f t, & 0 \leq t < T, \\ \pi \Delta f T, & t \geq T. \end{cases} \quad (2.26)$$

Колебания, соответствующие соседним (разнесенным на Δf) частотам манипуляции, за символьный интервал «разбегаются» по фазе на $2\pi\Delta f T$ радиан. Эта величина, нормированная к 2π , называется *индексом частотной модуляции* и обозначается буквой h :

$$h = \Delta f T.$$

Начальные фазы соседних символов при использовании ЧМНФ связаны друг с другом следующим образом:

$$\varphi_k = \varphi_{k-1} + 2\pi f_{k-1} T,$$

т. е. к начальной фазе предыдущего символа добавляется набег фазы, произошедший при передаче этого символа.

На практике используются сигналы ЧМНФ не только описанного вида (2.26), но и такие, у которых форма частотных посылок является непрямоугольной (в частности, *сглаженной*) и/или посылка не укладывается по длительности в символьный интервал. Использование таких сигналов позволяет еще больше сгладить форму сигнала и, следовательно, сделать его спектр еще более компактным. Однако оптимальный прием таких сигналов является достаточно сложной задачей (см. далее 2.2.7).

Среди возможных форм сглаженной ЧМНФ наибольшее практическое распространение получила *гауссова ЧМНФ* (англ. термин — Gaussian FSK (GFSK), при индексе модуляции $h = 1/2$ — Gaussian MSK (GMSK)), у которой частотная посылка $q(t)$ представляет собой свертку прямоугольного импульса и гауссовой функции. Такая разновидность ЧМНФ используется, в частности, в системах Bluetooth и GSM.

2.2.4. Общие методы приема

Рассмотрим методы приема, которые могут быть использованы для любого из описанных вариантов частотной манипуляции. В ситуации, когда решение о переданном символе принимается на каждом символьном интервале изолированно, данные методы являются оптимальными. Алгоритм приема ЧМНФ-сигнала, учитывающий непрерывный характер фазовой функции, будет рассмотрен отдельно в 2.2.7.

Как и в случае линейной модуляции, задача принятия решения о переданном символе — это задача различения M сигналов. Алгоритм ее оптимального решения зависит от того, известны начальные фазы сигнальных посылок или нет. Так как частота является неэнергетическим параметром, вместо минимизации евклидова расстояния в оптимальных алгоритмах можно максимизировать корреляцию принятого и образцовых сигналов.

При известных начальных фазах осуществляется *когерентный прием* и оптимальный (по методу максимума правдоподобия) приемник представляет собой M -канальный коррелятор (рис. 2.13).

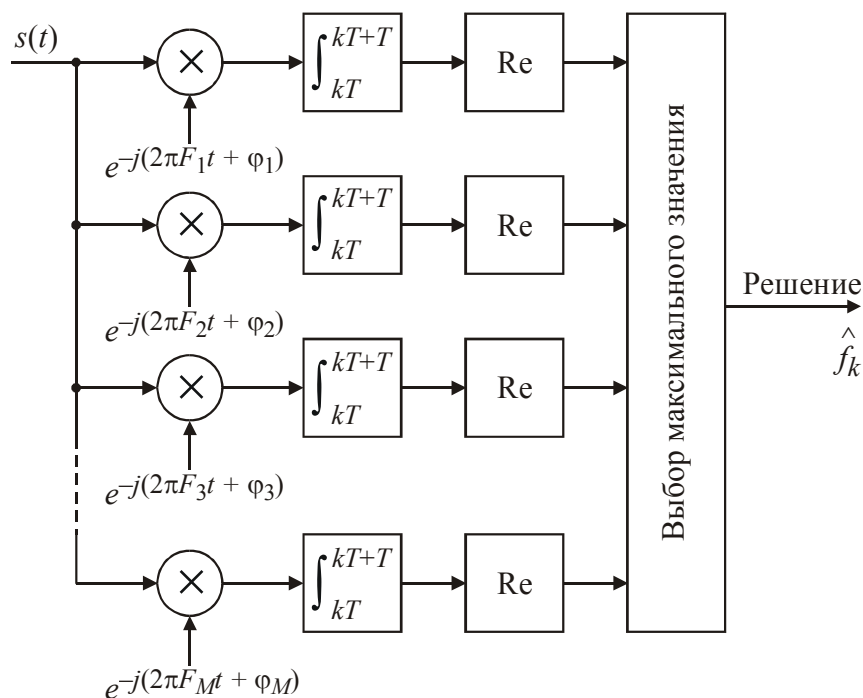


Рис. 2.13. Когерентный приемник ЧМ-сигнала

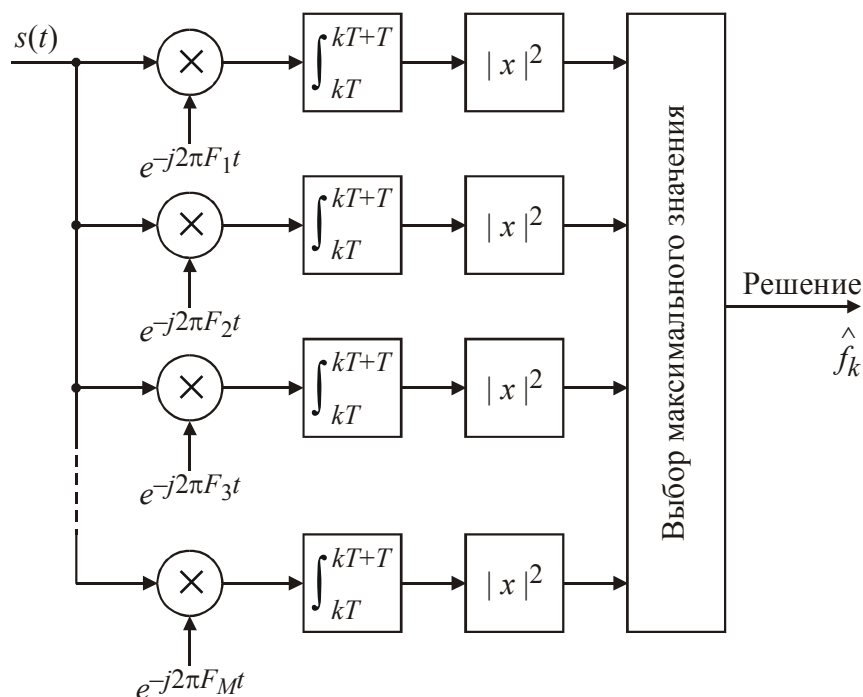


Рис. 2.14. Некогерентный приемник ЧМ-сигнала

При обработке комплексной огибающей вычисляются комплексные корреляции принятого сигнала со всеми возможными образцами, которые

имеют вид комплексных гармонических колебаний $\exp(j2\pi F_m t + j\varphi_m)$, после чего выбирается тот сигнал, для которого *вещественная часть* полученного результата максимальна.

Если начальные фазы рассматриваются как случайные величины с равномерным распределением в пределах $0 \dots 2\pi$, осуществляется *некогерентный прием* и оптимальный (по методу максимума правдоподобия) приемник представляет собой M -канальный некогерентный коррелятор, в котором, в отличие от когерентной схемы (рис. 2.13), вместо вещественных частей сравниваются *модули* комплексных корреляций, или, что эквивалентно и более удобно при практической реализации, их *квадраты* (рис. 2.14).

Еще одним отличием данной схемы от когерентного коррелятора является то, что в опорных сигналах нет необходимости учитывать начальные фазы.

2.2.5. Помехоустойчивость

Когерентный прием. При бинарной передаче (ЧМ-2) для произвольной пары сигналов $s_1(t)$ и $s_2(t)$ вероятность битовой ошибки определяется следующим образом:

$$P_b = P_{\text{sym}} = Q\left(\sqrt{\frac{E_b(1 - \text{Re } \dot{r})}{N_0}}\right), \quad (2.27)$$

где \dot{r} — комплексный коэффициент корреляции между двумя используемыми сигналами:

$$\dot{r} = \frac{1}{E_b} \int_0^T \dot{s}_1(t) \dot{s}_2^*(t) dt.$$

В случае ортогональных сигналов ($\text{Re } \dot{r} = 0$) формула (2.27) упрощается и принимает следующий вид:

$$P_b = P_{\text{sym}} = Q\left(\sqrt{\frac{E_b}{N_0}}\right). \quad (2.28)$$

Сравнение формул (2.28) и (2.11) показывает, что при бинарной передаче и когерентном приеме переход от пары противоположных сигналов (модуляция ФМ-2) к паре ортогональных сигналов дает двукратный энергетический проигрыш, т. е. потери составляют примерно 3 дБ.

Кривая помехоустойчивости для когерентного приема ортогональной модуляции ЧМ-2 была показана на рис. 2.5 штриховой линией.

В случае многопозиционной ортогональной ЧМ вероятность символьной ошибки выражается в виде интеграла, который нельзя представить в виде замкнутого аналитического выражения [2, 5.2.2]:

$$P_{\text{sym}} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(1 - (1 - Q(y))^{M-1}\right) \exp\left(-\frac{1}{2}\left(y - \sqrt{\frac{2E_{\text{sym}}}{N_0}}\right)^2\right) dy.$$

Чтобы упростить расчеты, воспользуемся аддитивной верхней границей для вероятности объединения событий перепутывания отдельных пар сигналов:

$$P_{\text{sym}} \leq (M-1)Q\left(\sqrt{\frac{E_{\text{sym}}}{N_0}}\right) = (M-1)Q\left(\sqrt{\frac{mE_b}{N_0}}\right).$$

При использовании ортогонального ансамбля вероятности перепутывания всех пар сигналов одинаковы, поэтому, независимо от способа назначения m -битовых комбинаций частотам манипуляции, вероятности битовой и символьной ошибок связаны следующим образом. Пусть, для определенности, передается символ, соответствующий битовой комбинации из одних нулей. Для этого символа возможно $M-1$ вариантов символьной ошибки. Всего в этих ошибочных символах содержится $mM/2$ ошибочных бит (набор из M комбинаций по m бит содержит mM бит, ровно половина из них имеет единичное значение, а правильный символ единичных бит не содержит). Так как вероятности перепутывания всех пар символов при использовании ортогональных сигналов одинаковы, то, при возникновении символьной ошибки, в среднем в символе искажается $mM/2/(M-1)$ бит. Так как символ содержит m бит, то коэффициент, связывающий вероятности символьной и битовой ошибок, будет равен $M/2/(M-1)$. Таким образом, получаем следующую формулу:

$$P_b = \frac{2^{m-1}}{2^m - 1} P_{\text{sym}} = \frac{M}{2(M-1)} P_{\text{sym}}. \quad (2.29)$$

Некогерентный прием. При бинарной передаче (ЧМ-2) для произвольной пары сигналов вероятность битовой ошибки рассчитывается по весьма громоздким формулам (см., например, [2, прил. В]). В частном случае ортогональных сигналов ($|\dot{r}| = 0$) общая формула сильно упрощается и принимает следующий вид:

$$P_b = \frac{1}{2} \exp\left(-\frac{E_b}{2N_0}\right). \quad (2.30)$$

Кривая помехоустойчивости для некогерентного приема ортогональной модуляции ЧМ-2 была показана на рис. 2.5 штрихпунктирной линией.

В случае многопозиционной ортогональной ЧМ вероятность ошибки при некогерентном приеме определяется следующим выражением [1, 4.9.2], [2, 5.4.3]:

$$P_{\text{sym}} = \frac{1}{M} \exp\left(-\frac{E_{\text{sym}}}{N_0}\right) \sum_{n=2}^M (-1)^n C_M^n \exp\left(\frac{E_{\text{sym}}}{nN_0}\right),$$

где C_M^n — число сочетаний из M по n (биномиальный коэффициент):

$$C_M^n = \frac{M!}{(M-n)!n!} = \frac{M(M-1)(M-2)\cdots(M-n+1)}{n!}.$$

При некогерентном приеме, как и в когерентном случае, ортогональный ансамбль обеспечивает равенство вероятностей перепутывания всех пар сигналов. Поэтому при некогерентном приеме вероятности битовой и символьной ошибок также связаны между собой формулой (2.29).

2.2.6. Условие ортогональности и ширина спектра

Как видно из приведенных в 2.2.5 формул, описывающих помехоустойчивость цифровой частотной модуляции, важнейшим параметром, влияющим на вероятность ошибки, является коэффициент корреляции между используемыми сигналами. Рассмотрим, чему равен этот коэффициент при использовании в качестве сигнальных посылок пары гармонических сигналов с разными частотами.

Итак, комплексные огибающие рассматриваемых сигналов равны:

$$s_1(t) = e^{j(2\pi f_1 t + \varphi_1)}, \quad s_2(t) = e^{j(2\pi f_2 t + \varphi_2)},$$

где $f_{1,2}$ — частоты сигналов; $\varphi_{1,2}$ — их начальные фазы. Коэффициент корреляции между этими двумя сигналами

$$\dot{r} = \frac{1}{T} \int_0^T s_1(t) s_2^*(t) dt = e^{j\Delta\varphi} e^{j\pi\Delta f T} \frac{\sin(\pi\Delta f T)}{\pi\Delta f T}. \quad (2.31)$$

Здесь $\Delta f = f_1 - f_2$ — расстояние между используемыми частотами; $\Delta\varphi = \varphi_1 - \varphi_2$ — разность начальных фаз сигналов.

При некогерентном приеме начальная фаза сигнала рассматривается как случайная величина, так что имеет значение только модуль полученного коэффициента корреляции:

$$|\dot{r}| = \frac{\sin(\pi\Delta f T)}{\pi\Delta f T}. \quad (2.32)$$

Наилучшая помехоустойчивость в случае некогерентного приема будет достигнута при нулевом значении коэффициента корреляции. Из формулы (2.32) получается следующее условие ортогональности сигналов при некогерентном приеме:

$$\Delta f = \frac{k}{T}, \quad k = \pm 1, \pm 2, \dots$$

Таким образом, минимальный разнос частот манипуляции, обеспечивающий ортогональность сигналов при произвольных начальных фазах, равен символьной скорости: $\Delta f = 1/T = R_{\text{sym}}$ (индекс модуляции h при этом равен единице). Именно эта величина используется в качестве оценки ширины спектра ЧМ-сигналов. Считая, что каждая из возможных частот манипуляции также занимает полосу шириной R_{sym} , получаем, что ширина спектра ЧМ-сигнала будет равна $M R_{\text{sym}} = 2^m R_{\text{sym}}$. Из этой формулы видно, что при увеличении числа бит в символе ширина спектра ЧМ-сигнала экспоненциально возрастает (в отличие от сигналов с линейной модуляцией, где ширина спектра от количества бит в символе не зависела).

Спектральная эффективность ЧМ-сигналов при использовании данной меры ширины спектра

$$(R/W)_{\text{ЧМ}} = (m R_{\text{sym}}) / (2^m R_{\text{sym}}) = m / 2^m.$$

При $m = 1$ и $m = 2$ спектральная эффективность одинакова и равна 0,5 (бит/с)/Гц:

$$1 / 2^1 = 2 / 2^2 = 0,5.$$

При дальнейшем увеличении числа бит в символе спектральная эффективность падает:

$$m = 3: \quad (R/W)_{\text{ЧМ}} = 3 / 2^3 = 0,375,$$

$$m = 4: \quad (R/W)_{\text{ЧМ}} = 4 / 2^4 = 0,25,$$

$$m = 5: \quad (R/W)_{\text{ЧМ}} = 5 / 2^5 = 0,15625 \text{ и т. д.}$$

При когерентном приеме помехоустойчивость определяется вещественной частью коэффициента корреляции (2.31):

$$\text{Re } \dot{r} = \cos(\pi \Delta f T + \Delta \varphi) \frac{\sin(\pi \Delta f T)}{\pi \Delta f T}. \quad (2.33)$$

Из (2.33) видно, что при когерентном приеме, в принципе, ортогональность сигналов может быть достигнута при произвольном (даже нулевом) расстоянии между используемыми частотами — за счет соответствующего выбора разности начальных фаз $\Delta \varphi$. Однако на практике когерентный прием (т. е. знание начальных фаз сигнальных посылок) можно реализовать разумными средствами, только сформировав ЧМ-сигнал с непрерывной фазой. При этом $\Delta \varphi = 0$ и формула (2.33) принимает вид

$$\text{Re } \dot{r} = \cos(\pi \Delta f T) \frac{\sin(\pi \Delta f T)}{\pi \Delta f T} = \frac{\sin(2\pi \Delta f T)}{2\pi \Delta f T}.$$

Отсюда видно, что в данном случае ортогональность сигналов достигается при вдвое меньшем разnose частот:

$$\Delta f = \frac{k}{2T}, \quad k = \pm 1, \pm 2, \dots$$

Минимальный сдвиг $\Delta f = 1/(2T) = R_{\text{sym}}/2$, что соответствует индексу модуляции $h = 1/2$. ЧМНФ с таким значением индекса называется *минимальной частотной манипуляцией* (МЧМ), или *частотной манипуляцией с минимальным сдвигом* (ЧММС) (англ. термин — Minimum Shift Keying (MSK)).

2.2.7. Прием частотно-манипулированных сигналов с непрерывной фазой

ЧМНФ является модуляцией *с памятью*, так как сигнал на каждом символьном интервале зависит не только от передаваемого в данный момент символа, но и от предыстории работы (т. е. от начальной фазы, накопившейся к началу символа). В связи с этим для организации оптимального приема такого сигнала необходимо рассматривать его *в целом*, а не анализировать отдельные символьные интервалы изолированно.

Алгоритм приема сигнала с памятью, реализующий нахождение последовательности символов по методу максимального правдоподобия, называется алгоритмом максимально правдоподобной оценки последовательности (англ. термин — Maximum Likelihood Sequence Estimation (MLSE)).

При приеме сигнала в белом гауссовом шуме для получения МП-оценки необходимо найти последовательность символов $\{a_k\}$, минимизирующую евклидово расстояние между соответствующим ЧМНФ-сигналом и принятой зашумленной реализацией. Однако ЧМНФ-сигнал является сигналом с постоянной огибающей, поэтому энергия любого его фрагмента не зависит от передаваемой информации. Это дает возможность вместо минимизации евклидова расстояния максимизировать корреляцию между возможными ЧМНФ-сигналами и принятой зашумленной реализацией. Дополнительным преимуществом при этом является отсутствие необходимости учитывать уровень принимаемого сигнала. Используя формулу (2.25) для записи фазовой функции ЧМНФ-сигнала, правило приема сигнала можно выразить так:

$$\int_0^{NT} \operatorname{Re} \left(\dot{s}(t) \exp \left(-j \sum_k a_k \phi(t - kT) \right) \right) dt \rightarrow \max_{\{a_k\}}. \quad (2.34)$$

Далее следует заметить, что если индекс модуляции h может быть представлен в виде рациональной дроби $h = M_h/P_h$, где M_h и P_h — взаимно простые натуральные числа, то число возможных начальных фаз сигнальных посылок оказывается конечным и равным P_h , если M_h четно, и $2P_h$, если M_h нечетно (в любом случае набор возможных начальных фаз равномерно распределен в пределах полной окружности). Это следует из того, что на протяжении каждого символьного интервала фаза сигнала изменяется на $\pm\pi h$ и, кроме того, она может быть определена только в пределах «одного оборота», т. е. с точностью до смещения, равного 2π . Таким образом, можно изобразить набор фазовых состояний сигнала и возможных переходов между ними. Пример такого изображения приведен на рис. 2.15 для индекса модуляции

$h = 1/2$ (МЧМ). В данном случае $M_h = 1$ нечетно, поэтому имеется $2P_h = 4$ фазовых состояний.

Таким образом, согласно критерию (2.34), необходимо найти на регулярной решетке, составленной из звеньев вида рис. 2.15, путь, максимизирующий корреляцию (2.34). Для решения именно такой задачи (задачи нахождения оптимального пути на заданной решетке) предназначен алгоритм Витерби, в обобщенном виде рассмотренный в 1.6.

В данном случае элементы обобщенного алгоритма конкретизируются следующим образом:

- «города» — это фазовые состояния сигнала (рис. 2.15);
- «дороги» — это возможные переходы между фазовыми состояниями (рис. 2.15);
- «расстояния» — это вещественные части комплексных корреляций между фрагментом принятой зашумленной реализации сигнала и сигналами, соответствующими возможным переходам между фазовыми состояниями (см. (2.34)). Так как необходимо добиться *максимального* значения критерия качества (2.34), при отборе «выживших» путей также необходимо искать не минимум, а максимум.

При рассмотрении алгоритма Витерби в обобщенном виде решетка была проанализирована целиком, и только потом при обратном просмотре был найден оптимальный путь. Для использования в связных системах такой подход во многих случаях неприемлем, так как решения о переданных символах необходимо выдавать, не дожидаясь окончания передачи всего сообщения. Просматривая набор выживших путей от текущего момента в прошлое, можно заметить, что через несколько символьных интервалов все выжившие пути сливаются друг с другом, образуя один общий путь. На этом участке окончательное решение уже определено, и оно не может измениться в процессе дальнейшей работы алгоритма.

Однако искать общий начальный фрагмент всех выживших путей и выдавать выходные решения по мере формирования этого фрагмента — вариант, неудобный для практической реализации. Дело в том, что глубина «ветвящейся» части набора выживших путей меняется во времени, поэтому выдача решений при указанном подходе будет происходить неравномерно, а для хранения анализируемой части решетки при реализации алгоритма потребуется переменное количество памяти. Поэтому на практике используется выдача решений с фиксированной задержкой. Эта задержка выбирается таким образом, чтобы выжившие пути сливались все вместе с достаточно

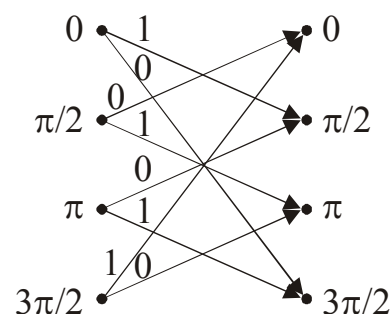


Рис. 2.15. Решетка переходов для ЧМНФ-сигнала с индексом модуляции $h = 1/2$

большой вероятностью. Многочисленные исследования показали [2, 5.1.4], что достаточной задержкой является пятикратная глубина памяти, имеющейся в принимаемом сигнале (под памятью понимается число предыдущих символов, от которых зависит формируемый в данный момент сигнал). Для ЧМНФ-сигнала глубина памяти равна длительности используемой частотной посылки $q(t)$. Таким образом, в случае несглаженной ЧМНФ достаточно просматривать последние 5 символьных интервалов и выдавать решения с задержкой, равной пяти символам.

Данный алгоритм приема может быть использован и для тех разновидностей ЧМНФ, у которых длительность частотной посылки $q(t)$ превышает длительность символа T . Число состояний в данном случае увеличивается, так как к фазовым добавляются еще и так называемые *корреляционные* состояния, позволяющие учесть тот факт, что фазовые переходы от отдельных символов накладываются друг на друга, так что фазовая функция сигнала на каждом символьном интервале зависит, кроме начальной фазы, не только от текущего передаваемого символа, но и от нескольких предыдущих.

2.3. Эффективность некодированной модуляции

На рис. 2.16 показаны достижимые соотношения между помехоустойчивостью и спектральной эффективностью для реальных видов некодированной модуляции. Требуемое отношение E_b/N_0 измерялось по уровню вероятности

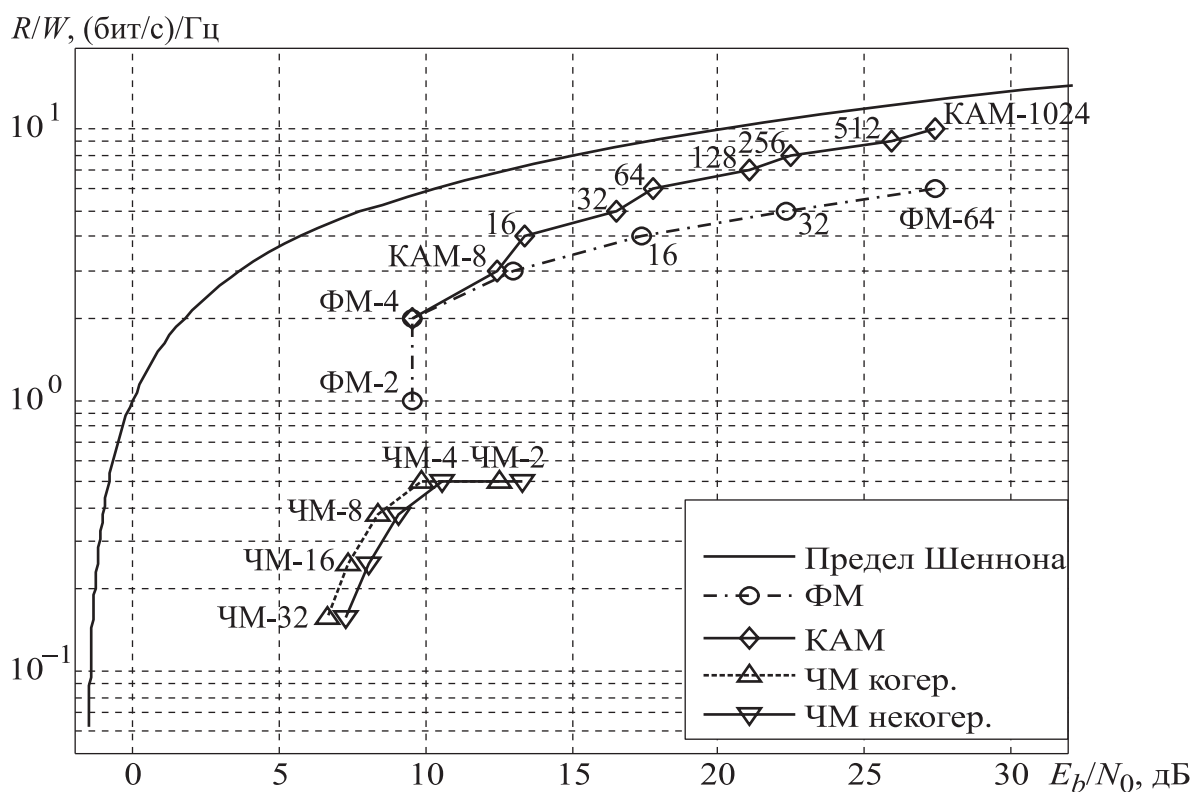


Рис. 2.16. Достижимые соотношения между помехоустойчивостью и спектральной эффективностью для некодированной модуляции

ошибки 10^{-5} . Данная вероятность была принята в качестве «практического критерия безошибочной передачи».

Из графика видно, что среди сигналов с линейной модуляцией наилучшее сочетание «помехоустойчивость — спектральная эффективность» обеспечивают КАМ-сигналы. Для всех сигналов с линейной модуляцией с ростом числа точек в сигнальном созвездии увеличивается спектральная эффективность и ухудшается помехоустойчивость. Таким образом, данные сигналы могут быть эффективно использованы в условиях большого отношения сигнал/шум при ограничениях на ширину доступной полосы частот.

ЧМ-сигналы, как показывает рис. 2.16, сочетают высокую помехоустойчивость и низкую спектральную эффективность. Таким образом, данный вид модуляции целесообразно использовать в ситуации, когда отношение сигнал/шум мало и нет серьезных ограничений на ширину спектра сигнала.

Что касается расстояния до границы Шеннона, рис. 2.16 демонстрирует, что для некодированной модуляции оно весьма велико и, в зависимости от конкретного вида модуляции, составляет 7...10 дБ.

3. МЕЖСИМВОЛЬНАЯ ИНТЕРФЕРЕНЦИЯ

До сих пор при рассмотрении алгоритмов приема сигналов с цифровой модуляцией предполагалось наличие АБГШ-канала, не искажающего форму полезного сигнала, а лишь добавляющего к нему белый шум. Амплитудно-частотная характеристика (АЧХ) такого канала является константой, фазочастотная характеристика (ФЧХ) линейно зависит от частоты, а импульсная характеристика (ИХ) представляет собой дельта-функцию. Однако характеристики реальных каналов связи отличаются от этих идеализированных предположений. Так, в проводных линиях связи частотная характеристика является неравномерной (затухание сигнала увеличивается с ростом частоты), а в радиоканале может наблюдаться многолучевое распространение радиоволн (сигнал может попадать от передающей антенны к приемной по нескольким путям; подробнее этот вопрос будет рассмотрен в разд. 4).

Во всех этих случаях условия отсутствия межсимвольной интерференции (2.17) и (2.18) оказываются нарушенными и вклады от отдельных символов, накладываясь друг на друга после согласованной фильтрации, будут снижать помехоустойчивость приема. Это явление, как уже отмечалось, называется межсимвольной интерференцией. В данном разделе будут описаны способы приема сигнала при наличии МСИ, характеристики которой постоянны во времени. Рассматриваемые алгоритмы относятся только к линейным видам модуляции.

3.1. Математическая модель канала с МСИ

Пусть импульсная характеристика канала связи равна $h(t)$, тогда после прохождения через канал связи и добавления шума комплексная огибающая сигнала с линейной модуляцией (2.1) будет иметь вид

$$s_h(t) = \sum_{m=-\infty}^{\infty} a_m g_h(t - mT) + n(t),$$

где $g_h(t)$ — сигнальная посылка, искаженная каналом связи, т. е. свертка исходной сигнальной посылки $g(t)$ и импульсной характеристики канала $h(t)$:

$$g_h(t) = \int_{-\infty}^{\infty} g(t') h(t - t') dt'.$$

После пропускания принятого сигнала через согласованный с сигнальной посылкой фильтр (см. разд. 2) получим сигнал

$$y(t) = \int_{-\infty}^{\infty} s_h(t') g(-(t - t')) dt' = \sum_{m=-\infty}^{\infty} a_m B_{gh}(t - mT) + \int_{-\infty}^{\infty} n(t') g(t' - t) dt',$$

где $B_{gh}(t)$ — свертка ИХ канала связи и КФ сигнальной посылки:

$$B_{gh}(t) = \int_{-\infty}^{\infty} h(t')B_g(t-t')dt'.$$

Дискретизируя сигнал $y(t)$ в моменты kT , получаем отсчеты следующего вида:

$$r_k = y(kT) = \sum_{m=-\infty}^{\infty} a_m B_{gh}((k-m)T) + \int_{-\infty}^{\infty} n(t')g(t'-kT)dt' = \sum_{m=-\infty}^{\infty} a_m h_{k-m} + n_k,$$

где h_k — отсчеты эквивалентной дискретной ИХ канала связи с МСИ:

$$h_k = B_{gh}(kT),$$

а n_k — отсчеты шума. На преобразования шума МСИ никак не влияет, поэтому отсчеты шума после приемного фильтра и дискретизации будут, как и ранее (см. разд. 2), представлять собой дискретный белый шум.

Таким образом, окончательно математическую модель сигнала с МСИ можно записать следующим образом:

$$\dot{r}_k = \sum_{m=-\infty}^{\infty} \dot{a}_m \dot{h}_{k-m} + \dot{n}_k, \quad (3.1)$$

где \dot{r}_k — отсчеты принимаемого сигнала после СФ; \dot{h}_k — отсчеты эквивалентной импульсной характеристики канала связи; \dot{n}_k — отсчеты АБГШ.

При дальнейшем рассмотрении алгоритмов приема сигнала, искаженного МСИ, предполагается, что ИХ канала известна. На практике оценка ИХ канала может быть получена благодаря тому, что в передаваемый поток данных включаются специальные фрагменты, генерируемые по закону, известному на приемной стороне, и специально предназначенные для подстройки приемных алгоритмов под характеристики конкретного канала связи.

3.2. Оптимальный алгоритм приема сигнала (алгоритм Витерби)

Для оптимального (по методу МП) приема сигнала при наличии МСИ и знании ИХ канала необходимо найти такую последовательность символов $\{\dot{a}_k\}$, которая бы минимизировала евклидово расстояние между сигналом с МСИ и принятой последовательностью отсчетов:

$$\sum_{k=1}^K \left| \dot{r}_k - \sum_{m=-\infty}^{\infty} \dot{h}_m \dot{a}_{k-m} \right|^2 \rightarrow \min_{\{\dot{a}_k\}}. \quad (3.2)$$

Реализовать поиск этой последовательности можно с помощью алгоритма Витерби. Рассмотрим, как в данном случае строится решетка состояний и переходов между ними («городов» и «дорог» в обобщенном представлении алгоритма Витерби, рассмотренном в 1.6).

В реальных ситуациях длина эквивалентной импульсной характеристики канала $\{\dot{h}_k\}$ является конечной, пусть она равна L ($\dot{h}_k = 0$ при $k < 0$ и при $k \geq L$). Тогда, согласно (3.1), каждый отсчет \dot{r}_k зависит от L переданных символов: $\{a_{k-L+1}, \dots, a_k\}$. При переходе к следующему отсчету \dot{r}_{k+1} ($L-1$) из этих символов ($\{a_{k-L+2}, \dots, a_{k+1}\}$) «наследуются» — исчезает зависимость лишь от самого «старого» символа a_{k-L+1} , вместо этого появляется зависимость от очередного символа a_{k+1} . Эта связь между отсчетами принимаемого сигнала и позволяет осуществлять его последовательный прием (или прием *в целом*) с учетом имеющейся в сигнале МСИ. Английский термин для данного алгоритма — Maximum Likelihood Sequence Estimation (MLSE).

Итак, состоянием в данном случае будет являться последовательность значений *предыдущих* символов (в количестве $L-1$), от которых зависит отсчет принимаемого сигнала. Так как каждый символ может принимать одно из M возможных значений, общее число состояний равно M^{L-1} .

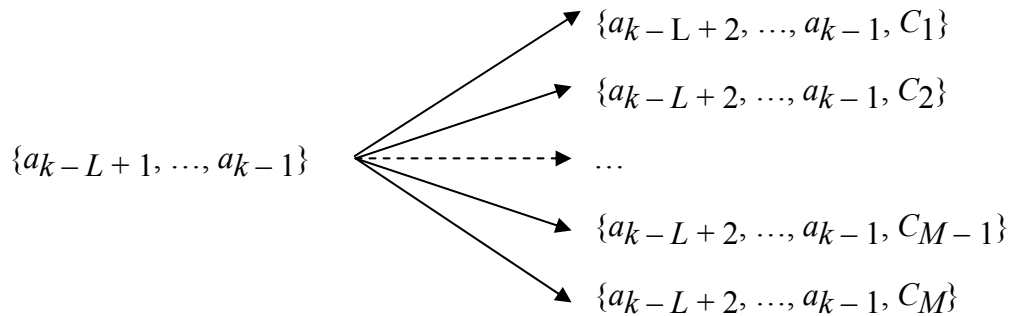


Рис. 3.1. Набор возможных переходов для каждого состояния решетки алгоритма Витерби

Из каждого конкретного состояния $\{a_{k-L+1}, \dots, a_{k-1}\}$ на следующем такте можно попасть не в любое состояние, а только в одно из M состояний, соответствующих последовательностям значений символов, первые $L-2$ из которых имеют вид $\{a_{k-L+2}, \dots, a_{k-1}\}$, а последний символ может принимать любое из M возможных значений C_1, C_2, \dots, C_M . Набор переходов, возможных из каждого состояния, в обобщенном виде показан на рис. 3.1.

В качестве примера на рис. 3.2 показана полная решетка переходов для четырехпозиционной модуляции ($M=4$) и трехэлементной импульсной характеристики канала связи ($L=3$). Число состояний в данном случае равно $M^{L-1} = 4^3 - 1 = 4^2 = 16$.

Каждому переходу из состояния в состояние соответствует комбинация из L символов $\{a_{k-L+1}, \dots, a_k\}$. Этой комбинации, в свою очередь, соответствует вполне определенное значение сигнала, согласно (3.1) равное

$$\dot{r}_{\{\dot{a}_{k-L+1}, \dots, \dot{a}_k\}} = \sum_{m=0}^{L-1} \dot{a}_{k-m} \dot{h}_m.$$

Квадрат евклидова расстояния между этим значением и принятым отсчетом сигнала \dot{r}_k , согласно (3.2), дает отдельное слагаемое в минимизируемой целевой функции («расстояние», согласно терминологии из 1.6).

Далее приведен пример использования алгоритма Витерби для приема сигнала при наличии МСИ (рис. 3.3).

Используется модуляция ФМ-2 ($M = 2$, $m = 1$), так что передаваемые символы могут принимать значения ± 1 . Канал связи имеет импульсную характеристику, состоящую из двух отсчетов ($L = 2$): $\{h_k\} = \{+1, -0,9\}$, так что принятые отсчеты, согласно (3.1), равны

$$r_k = a_k - 0,9a_{k-1} + n_k.$$

Набор возможных значений полезного сигнала представлен в табл. 3.1.

Число состояний алгоритма Витерби $M^{L-1} = 2^{2-1} = 2$. Одна ячейка решетки переходов показана на рис. 3.4. В начале стрелок переходов, выходящих из исходных состояний, подписаны значения вновь поступившего на вход символа (a_k). Значения сигнала из табл. 3.1 показаны в квадратных рамках поверх стрелок соответствующих переходов.

Рассмотрим подробно обработку двух первых отсчетов сигнала с помощью алгоритма Витерби (рис. 3.5).

Для каждого состояния алгоритм хранит *метрику* — квадрат евклидова расстояния между обработанным фрагментом сигнала и выбранным к текущему моменту оптимальным путем, заканчивающимся в данном состоянии. Исходно метрики для всех состояний равны нулю.

Первый зашумленный отсчет равен $+1,3$ (см. рис. 3.3). Для этого значения вычисляются квадраты разности со всеми возможными значениями сигнала, приведенными ранее в табл. 3.1 (на рис. 3.5 этот этап обозначен как «приращение метрики»).

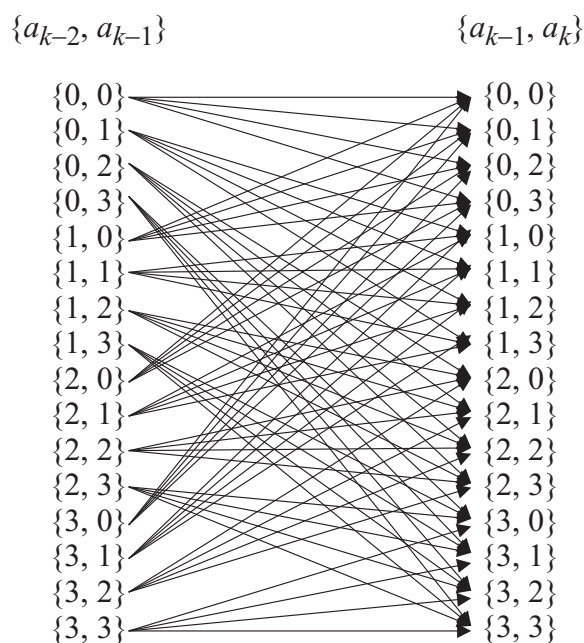


Рис. 3.2. Решетка алгоритма Витерби для случая $L = 3$ и $M = 4$

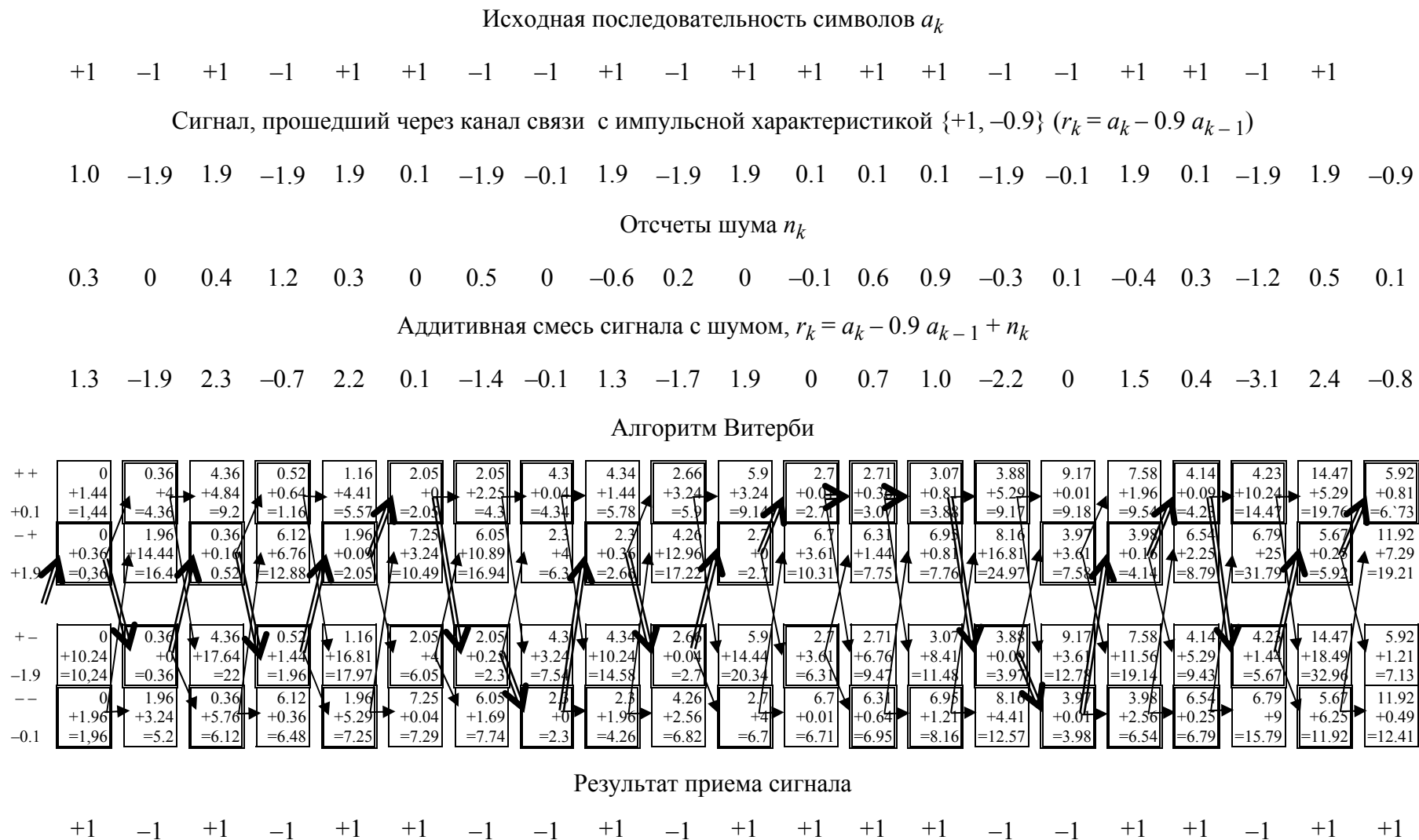


Рис. 3.3. Оптимальный прием сигнала с линейной модуляцией при наличии МСИ

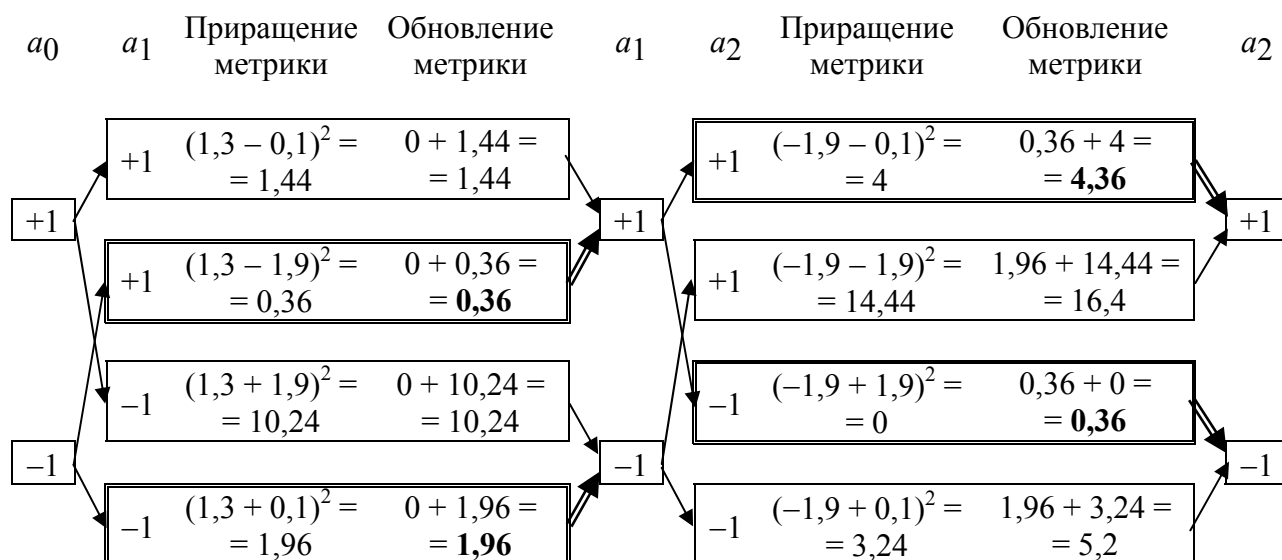


Рис. 3.5. Начало работы алгоритма Витерби для рассматриваемого примера

Табл. 3.1. Возможные значения полезного сигнала

a_{k-1}	a_k	$r_{\{a_{k-1}, a_k\}}$
+1	+1	$+1 - 0,9 = +0,1$
+1	-1	$-1 - 0,9 = -1,9$
-1	+1	$+1 + 0,9 = +1,9$
-1	-1	$-1 + 0,9 = -0,1$

Далее для каждого возможного перехода между состояниями (см. рис. 3.4) производится расчет обновленной метрики, представляющей собой сумму метрики исходного состояния и рассчитанного ранее приращения метрики, соответствующего данному переходу. На рис. 3.5 этот этап обозначен как «обновление метрики». При обработке первого отсчета сигнала все метрики исходных состояний равны нулю, так что обновленные метрики оказываются равны рассчитанному приращению.

После расчета обновленных метрик выполняется отбор выживших путей. Для этого сравниваются метрики всех путей, ведущих в одно и то же состояние, и выбирается в качестве выжившего тот из них, метрика которого минимальна. Это минимальное значение становится метрикой данного состояния.

В данном примере в состояние «+1» ведут 2 пути: путь, выходящий из состояния «+1» и имеющий метрику 1,44, и путь, выходящий из состояния «-1» и имеющий метрику 0,36. В результате сравнения метрик второй путь помечается как выживший (расчеты метрики для него обведены на рис. 3.5 двойной рамкой, а сам переход обозначен двойной стрелкой), а значение 0,36 становится метрикой состояния «+1» (это значение выделено на рис. 3.5 полужирным шрифтом).

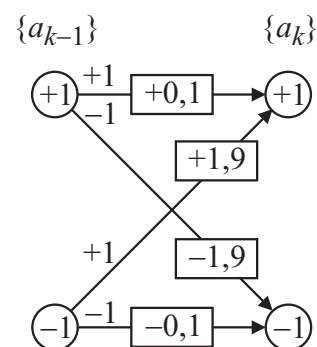


Рис. 3.4. Решетка алгоритма Витерби для рассматриваемого примера ($L = 2$ и $M = 2$)

В состояние «-1» также ведут 2 пути: из состояния «+1» (метрика пути 10,24) и из состояния «-1» (метрика пути 1,96). Метрика второго пути лучше, поэтому он помечается как выживший, а значение 1,96 становится метрикой состояния «-1» (выживший путь и итоговая метрика обозначены на рис. 3.5 так же, как и для состояния «+1»).

Далее аналогичным образом обрабатывается второй отсчет сигнала, равный -1,9: рассчитываются приращения метрик, затем вычисляются обновленные метрики (они уже не равны приращениям метрик, как это было для первого отсчета, так как метрики исходных состояний перестали быть нулевыми). В состояние «+1» ведут пути с метриками 4,36 и 16,4; выбирается первый из них. В состояние «-1» ведут пути с метриками 0,36 и 5,2; выбирается первый из них. Выбранные пути и итоговые метрики на рис. 3.5 обозначены так же, как для первого отсчета.

После обработки двух отсчетов сигнала уже видно, что первый декодированный символ (a_1) равен +1, так как через это состояние проходят оба выживших пути.

Далее аналогичным образом обрабатываются остальные отсчеты входного сигнала: рассчитываются приращения метрик, обновляются метрики, выбираются выжившие пути. Результаты расчетов показаны на рис. 3.3 в более кратком представлении, чем на рис. 3.5: расчет приращений метрик не показан, не показаны и сами состояния, приведены только вычисления для обновления метрик. Кроме того, на этом рисунке показаны только выжившие пути, а двойными стрелками выделен результат обратной трассировки (см. 1.6), т. е. искомый оптимальный путь.

Приведенный в нижней части рис. 3.3 результат приема сигнала показывает, что ошибок в данном случае не произошло.

Все изложенное в 2.2.7 о выдаче сформированных алгоритмом Витерби решений справедливо и в данном случае. Как правило, при реализации алгоритма хранится фрагмент решетки, имеющий фиксированную протяженность по времени, а решение выдается путем обратного просмотра этой решетки на фиксированную глубину. Экспериментально показано, что достаточной глубиной просмотра является пятикратная длительность импульсной характеристики канала связи [2].

На рис. 3.6 показаны характеристики помехоустойчивости приема сигнала, прошедшего через рассмотренный канал связи, соответствующие разным алгоритмам приема. Из графика, соответствующего алгоритму Витерби (сплошная линия), видно, что проигрыш случаю отсутствия МСИ (пунктирная кривая) невелик и составляет примерно 1 дБ.

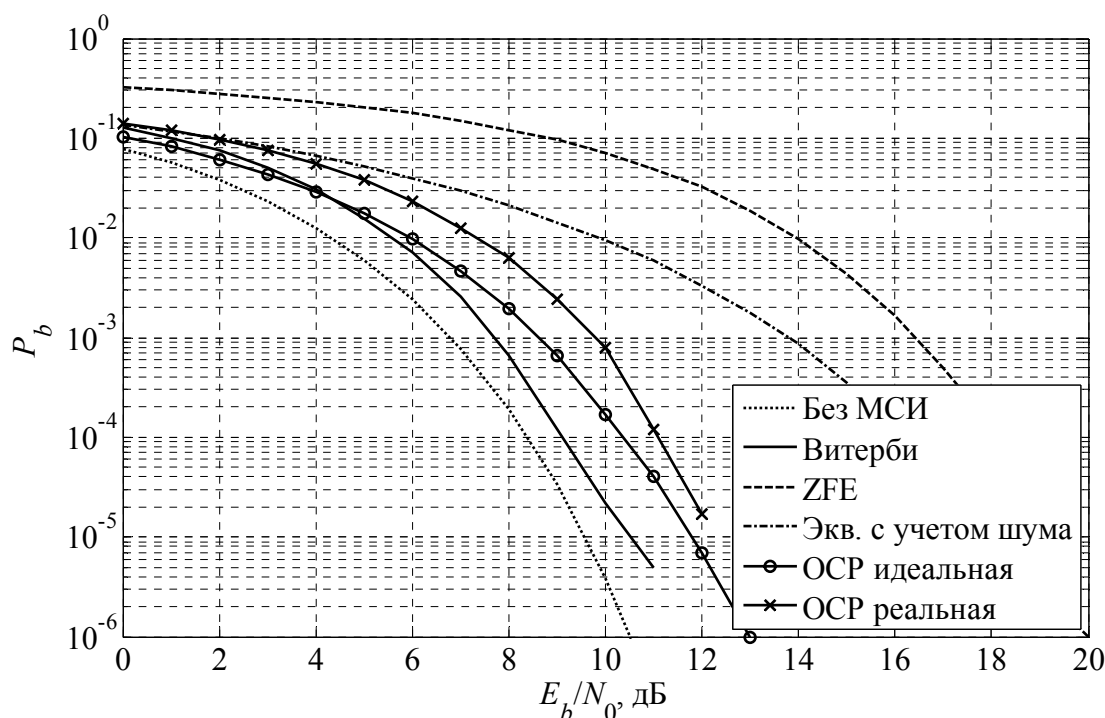


Рис. 3.6. Характеристики помехоустойчивости для различных способов приема сигнала с модуляцией ФМ-2, прошедшего через канал связи с ИХ $\{+1, -0,9\}$

Главным недостатком алгоритма Витерби является экспоненциальная зависимость числа состояний от длины ИХ канала связи. Это существенно ограничивает возможности практического применения данного оптимального алгоритма, поэтому на практике часто используются не являющиеся оптимальными, но практически реализуемые технические решения на основе фильтров. Такие подходы рассматриваются в данном разделе далее.

3.3. Оптимальный фильтр Винера

Фильтры, используемые для приема сигнала при наличии МСИ, реализуют оптимальную фильтрацию дискретного случайного процесса. Выведем общие формулы для расчета оптимального фильтра [5], а затем рассмотрим их применение к задаче подавления МСИ.

Замечание. Нужно понимать, что оптимальный фильтр, используемый при приеме сигнала, решает иную задачу, нежели рассмотренный ранее оптимальный алгоритм Витерби — вместо поиска наиболее правдоподобной последовательности символов он лишь пытается наилучшим образом очистить сигнал от МСИ и шума.

Для упрощения выкладок вывод производится для вещественных сигналов. Для более общего случая обработки комплексных сигналов полученный результат сохраняет свою форму, однако при реализации соответствующих расчетов необходимо помнить, что в определениях корреляционных функций

для комплексных сигналов один из сигналов подвергается комплексному сопряжению.

Говоря об оптимальной фильтрации сигнала, следует помнить, что данная задача становится осмысленной лишь после задания двух вещей — математической модели входного сигнала и оптимизируемого критерия качества. Тогда задача оптимальной фильтрации сводится к математической оптимизационной задаче, которая может быть решена аналитически либо численно.

Итак, пусть входной дискретный случайный сигнал $\{x(k)\}$ обрабатывается дискретным фильтром порядка N с коэффициентами $\{w_n\}$, $n = 0, 1, \dots, N$ (рис. 3.7). Выходной сигнал фильтра

$$y(k) = \sum_{n=0}^N w_n x(k-n).$$

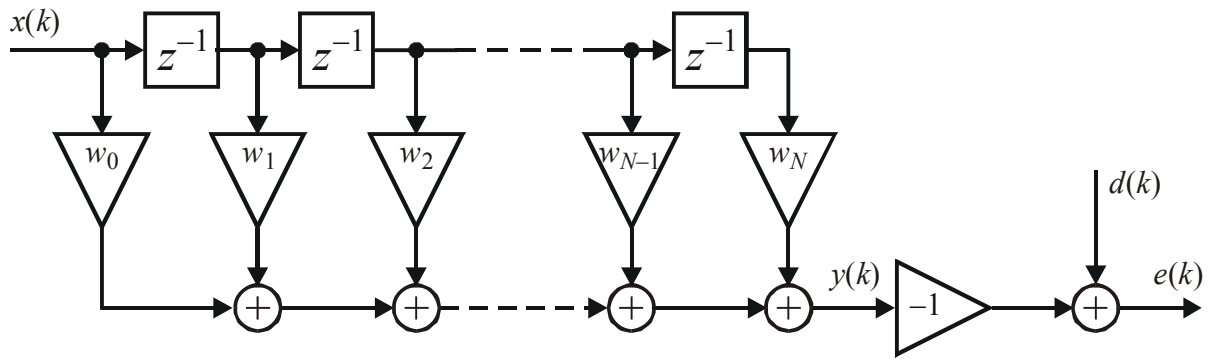


Рис. 3.7. Формирование сигнала ошибки

Замечание. Следует помнить о различии понятий «порядок фильтра» и «длина фильтра». Порядок фильтра — это число *задержанных* отсчетов сигнала, используемых при вычислениях. Длина фильтра на единицу больше, так как в вычислениях участвует также текущий (незадержанный) отсчет.

Кроме того, имеется образцовый (также случайный) сигнал $\{d(k)\}$. Ошибка воспроизведения образцового сигнала

$$e(k) = d(k) - y(k) = d(k) - \sum_{n=0}^N w_n x(k-n). \quad (3.3)$$

Необходимо найти такие коэффициенты фильтра $\{w_n\}$, которые обеспечивают максимальную близость выходного сигнала фильтра к образцовому, т. е. минимизируют ошибку $e(k)$. Но, поскольку $e(k)$ также является случайным процессом, в качестве меры ее значения разумно принять средний квадрат. Таким образом, оптимизируемая функция выглядит так:

$$J(\{w_n\}) = \overline{e^2(k)} \rightarrow \min.$$

Для решения поставленной задачи прежде всего перепишем (3.3) в матричном виде. Для этого обозначим вектор-столбец коэффициентов фильтра

как \mathbf{w} , а вектор-столбец содержимого линии задержки фильтра на k -м шаге как $\mathbf{x}(k)$:

$$\mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_N \end{bmatrix}, \quad \mathbf{x}(k) = \begin{bmatrix} x(k) \\ x(k-1) \\ \vdots \\ x(k-N) \end{bmatrix}.$$

С учетом этих обозначений (3.3) принимает следующий вид:

$$e(k) = d(k) - \mathbf{x}^T(k)\mathbf{w}.$$

Квадрат ошибки

$$\begin{aligned} e^2(k) &= \left(d(k) - \mathbf{x}^T(k)\mathbf{w} \right)^2 = d^2(k) - 2d(k)\mathbf{x}^T(k)\mathbf{w} + \left(\mathbf{x}^T(k)\mathbf{w} \right)^2 = \\ &= d^2(k) - 2d(k)\mathbf{x}^T(k)\mathbf{w} + \mathbf{w}^T \mathbf{x}(k)\mathbf{x}^T(k)\mathbf{w}. \end{aligned}$$

Статистически усредняя это выражение, получаем следующее:

$$J(\mathbf{w}) = \overline{e^2(k)} = \overline{d^2(k)} - 2\left(\overline{d(k)\mathbf{x}(k)}\right)^T \mathbf{w} + \mathbf{w}^T \overline{\mathbf{x}(k)\mathbf{x}^T(k)}\mathbf{w}. \quad (3.4)$$

Рассмотрим подробнее входящие в полученную формулу усредненные величины:

- $\overline{d^2(k)}$ — это средний квадрат образцового сигнала. Он представляет собой отдельное слагаемое, которое не зависит от коэффициентов фильтра и потому может быть отброшено (однако оно влияет на значение среднего квадрата ошибки, получаемое при оптимальных коэффициентах фильтра). Обозначим данное значение как σ_d^2 ;
- $\overline{d(k)\mathbf{x}(k)}$ — это вектор-столбец взаимных корреляций между k -м отсчетом образцового сигнала и содержимым линии задержки фильтра на k -м шаге. Будем считать случайные процессы $x(k)$ и $d(k)$ совместно стационарными, тогда вектор взаимных корреляций не зависит от номера шага k . В дальнейших выкладках этот вектор будет обозначен как \mathbf{p} :

$$\mathbf{p} = \begin{bmatrix} \overline{d(k)x(k)} \\ \overline{d(k)x(k-1)} \\ \vdots \\ \overline{d(k)x(k-N)} \end{bmatrix}; \quad (3.5)$$

- $\overline{\mathbf{x}(k)\mathbf{x}^T(k)}$ — это квадратная матрица размером $(N+1) \times (N+1)$, являющаяся корреляционной матрицей сигнала. Для стационарного случайного процесса корреляционная матрица имеет *теплицеву структуру*, т. е. вдоль

диагоналей матрицы стоят одинаковые значения, являющиеся значениями корреляционной функции:

$$\mathbf{R} = \begin{bmatrix} R_x(0) & R_x(1) & R_x(2) & \dots & R_x(N) \\ R_x(1) & R_x(0) & R_x(1) & \dots & R_x(N-1) \\ R_x(2) & R_x(1) & R_x(0) & \dots & R_x(N-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R_x(N) & R_x(N-1) & R_x(N-2) & \dots & R_x(0) \end{bmatrix}, \quad (3.6)$$

где $R_x(\Delta k) = \overline{x(k)x(k-\Delta k)}$ — КФ входного сигнала.

С учетом введенных обозначений (3.4) принимает следующий вид:

$$J(\mathbf{w}) = \sigma_d^2 - 2\mathbf{p}^T \mathbf{w} + \mathbf{w}^T \mathbf{R} \mathbf{w}. \quad (3.7)$$

Данное выражение представляет собой квадратичную форму относительно \mathbf{w} и потому при невырожденной матрице \mathbf{R} имеет единственный минимум, для нахождения которого необходимо приравнять к нулю вектор градиента:

$$\text{grad } J(\mathbf{w}) = -2\mathbf{p} + 2\mathbf{R}\mathbf{w} = \mathbf{0}.$$

Отсюда получаем равенство, называемое уравнением Винера—Хопфа (Wiener—Hopf equation):

$$\mathbf{R}\mathbf{w} = \mathbf{p}.$$

Умножив слева обе части равенства на обратную корреляционную матрицу \mathbf{R}^{-1} , получаем искомое решение для оптимальных коэффициентов фильтра:

$$\mathbf{w} = \mathbf{R}^{-1}\mathbf{p}. \quad (3.8)$$

Такой фильтр называется *фильтром Винера* (Wiener filter). Подстановка (3.8) в (3.7) дает минимально достижимую дисперсию сигнала ошибки:

$$\overline{e^2(k)}_{\min} = \sigma_d^2 - \mathbf{p}^T \mathbf{R}^{-1} \mathbf{p}.$$

Несложно также показать, что $\overline{e(k)y(k)} = 0$ и $\overline{e(k)\mathbf{x}(k)} = \mathbf{0}$, т. е. сигнал ошибки для фильтра Винера не коррелирован со входным и выходным сигналами фильтра.

3.4. Линейный эквалайзер

При использовании фильтра Винера для оптимальной фильтрации сигнала с МСИ входной сигнал — это сигнал с МСИ плюс АБГШ (см. (3.1)). В данном случае необходимо как можно лучше очистить принимаемый сигнал от шума и МСИ, поэтому образцовым сигналом должны являться информационные символы a_k , взятые с некоторой задержкой k_0 :

$$x(k) = \sum_{m=-\infty}^{\infty} \dot{a}_m \dot{h}_{k-m} + \dot{n}_k, \quad d(k) = a_{k-k_0}.$$

Как следует из (3.8), для расчета оптимального фильтра необходимо знать матрицу \mathbf{R} и вектор \mathbf{p} . Матрица \mathbf{R} , как видно из (3.6), заполнена значениями КФ входного сигнала. Рассчитаем эту КФ:

$$R_x(\Delta k) = \overline{x(k)x^*(k - \Delta k)} = \overline{\left(\sum_{m_1=-\infty}^{\infty} \dot{a}_{m_1} \dot{h}_{k-m_1} + \dot{n}_k \right) \left(\sum_{m_2=-\infty}^{\infty} \dot{a}_{m_2}^* \dot{h}_{k-\Delta k-m_2}^* + \dot{n}_{k-\Delta k}^* \right)}.$$

Раскрытие скобок в этом выражении, учет статистической независимости передаваемых символов и нулевого математического ожидания шума приводят к следующей формуле:

$$R_x(\Delta k) = \sigma_a^2 \sum_{k=-\infty}^{\infty} h_k h_{k-\Delta k}^* + \overline{n_k n_{k-\Delta k}^*}. \quad (3.9)$$

Входящая в это выражение сумма по k представляет собой КФ импульсной характеристики канала связи:

$$B_h(\Delta k) = \sum_{k=-\infty}^{\infty} h_k h_{k-\Delta k}^*. \quad (3.10)$$

Шум является белым, так что его КФ представляет собой одиночный дискретный импульс:

$$\overline{n_k n_{k-\Delta k}^*} = \sigma_n^2 \delta_{0, \Delta k} = \begin{cases} \sigma_n^2, & \Delta k = 0, \\ 0, & \Delta k \neq 0, \end{cases} \quad (3.11)$$

где δ — символ Кронекера.

С учетом (3.10) и (3.11) выражение (3.9) для КФ принимает окончательный вид:

$$R_x(\Delta k) = \sigma_a^2 B_h(\Delta k) + \sigma_n^2 \delta_{0, \Delta k}. \quad (3.12)$$

Для расчета n -го элемента вектора \mathbf{p} необходимо вычислить взаимную корреляцию между образцовым сигналом и задержанным на n отсчетов входным сигналом (см. (3.5)):

$$\begin{aligned} p_n &= \overline{x(k-n)d^*(k)} = \overline{\left(\sum_{m=-\infty}^{\infty} \dot{a}_m \dot{h}_{k-n-m} + \dot{n}_{k-n} \right) a_{k-k_0}^*} = \\ &= \sum_{m=-\infty}^{\infty} \overline{\dot{a}_m a_{k-k_0}^*} \dot{h}_{k-n-m} + \dot{n}_{k-n} a_{k-k_0}^*. \end{aligned} \quad (3.13)$$

В сумме по m , вследствие статистической независимости передаваемых символов, отлично от нуля лишь одно слагаемое (см. (2.2)), когда $m = k - k_0$. Кроме того, второе слагаемое (3.13) равно нулю благодаря нулевому математическому ожиданию шума и его независимости от полезного сигнала. Таким образом,

$$p_n = \sigma_a^2 h_{k-n-(k-k_0)} = \sigma_a^2 h_{k_0-n}, \quad n = 0, 1, \dots, N, \quad (3.14)$$

т. е. вектор \mathbf{p} содержит ИХ канала связи, зеркально перевернутую и сдвинутую в соответствии с задержкой фильтрации k_0 .

Теперь импульсная характеристика оптимального фильтра может быть рассчитана по (3.8). Рассчитанный указанным образом фильтр называется *линейным эквалайзером*, так как он представляет собой линейный фильтр и в определенной степени выравнивает частотную характеристику канала связи (см. далее рис. 3.9).

В качестве примера найдем матрицу \mathbf{R} и вектор \mathbf{p} для фильтра длины 8 (порядок фильтра $N=7$), рассчитанного на канал связи, использованный в 3.2: $\{h_k\} = \{+1, -0,9\}$. Задержку фильтрации примем равной пяти отсчетам, а отношение сигнал/шум равным 10 дБ (10 раз по мощности). КФ ИХ канала связи

$$B_h(0) = h_0^2 + h_1^2 = 1^2 + (-0,9)^2 = 1,81, \quad (3.15)$$

$$B_h(1) = B_h(-1) = h_0 h_1 = -0,9.$$

Для расчета дисперсии шума нужно заметить, что мощность полезного сигнала $\sigma_a^2 B_h(0) = 1,81 \sigma_a^2$. Дисперсию шума можно получить, разделив эту величину на отношение сигнал/шум по мощности:

$$\sigma_n^2 = \frac{\sigma_a^2 B_h(0)}{\text{SNR}} = \frac{1,81 \sigma_a^2}{10} = 0,181 \sigma_a^2. \quad (3.16)$$

Наконец, можно разделить матрицу \mathbf{R} и вектор \mathbf{p} на σ_a^2 , чтобы исключить из них этот незначительный параметр (согласно (3.8), этот общий масштаб не влияет на результат расчета). В итоге, воспользовавшись (3.12) и (3.14)–(3.16), получаем

$$\mathbf{R} = \begin{bmatrix} 1,991 & -0,9 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0,9 & 1,991 & -0,9 & 0 & 0 & 0 & 0 & 0 \\ 0 & -0,9 & 1,991 & -0,9 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0,9 & 1,991 & -0,9 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0,9 & 1,991 & -0,9 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0,9 & 1,991 & -0,9 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0,9 & 1,991 & -0,9 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0,9 & 1,991 \end{bmatrix}, \quad \mathbf{p} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -0,9 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

Результаты расчета показаны на рис. 3.8, а на рис. 3.9 приведены сквозные характеристики скорректированного канала связи (его ИХ представляет

собой свертку ИХ канала и фильтра, а его АЧХ — произведение АЧХ канала и фильтра).

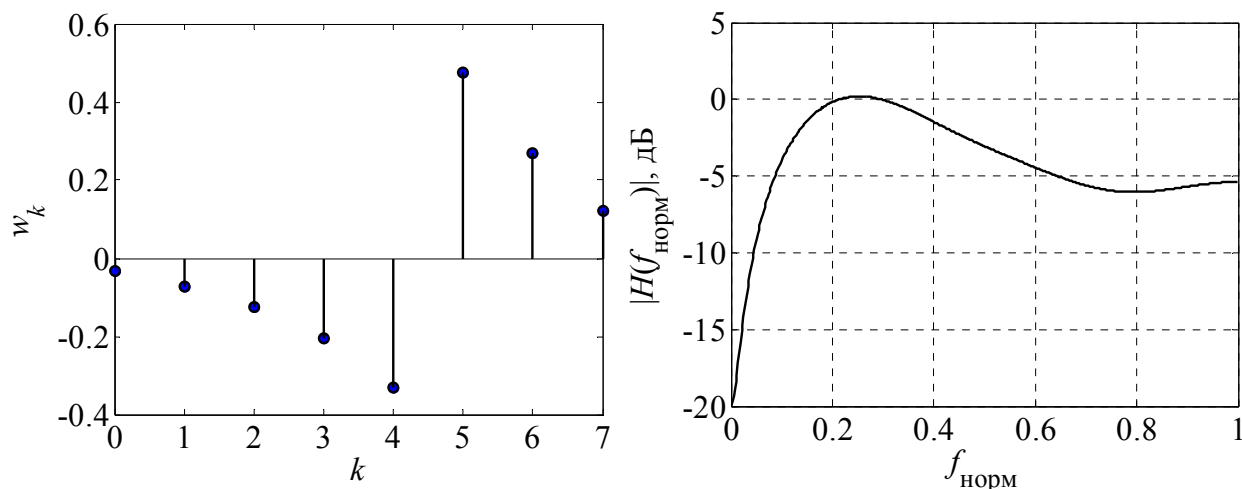


Рис. 3.8. ИХ (слева) и АЧХ (справа) линейного эквалайзера 7-го порядка (SNR = 10 дБ, задержка 5 отсчетов)

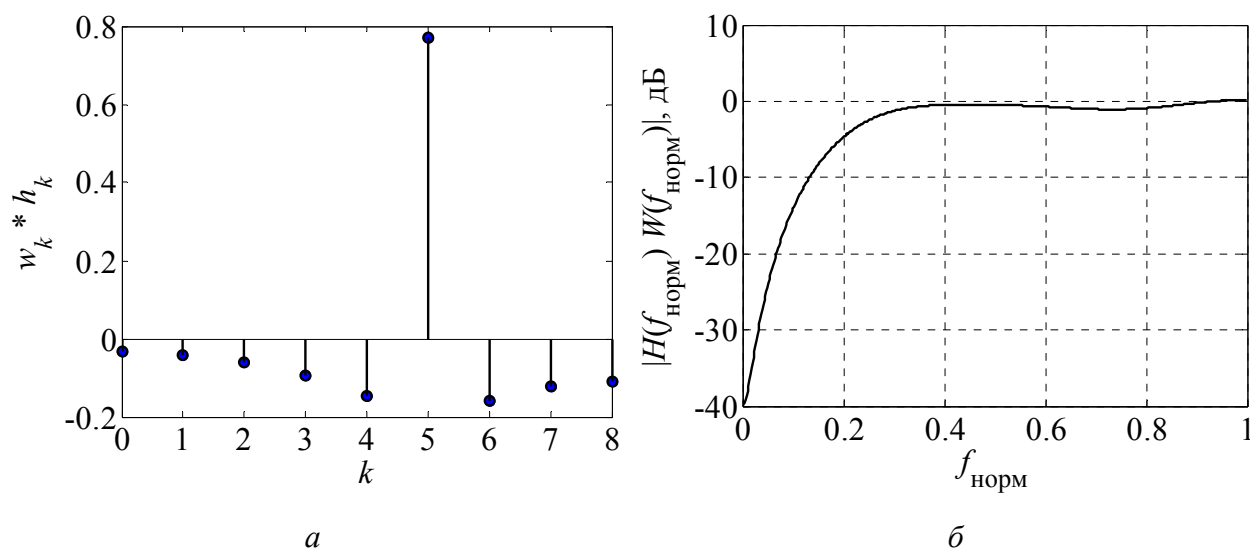


Рис. 3.9. Характеристики скорректированного канала связи: ИХ (а) и АЧХ (б)

Из рис. 3.9, а видно, что сквозная ИХ скорректированного канала связи представляет собой резко выраженный пик, горизонтальное положение которого соответствует заданной при расчете задержке фильтрации, а амплитуда близка к единице (она должна быть равна единице при идеальной компенсации МСИ). Справа и слева от этого пика ИХ имеет отсчеты небольшого уровня, которые дают остаточную МСИ.

Сквозная АЧХ скорректированного канала связи (рис. 3.9, б) близка к единице (0 дБ) на большей части частотной оси, за исключением области низких частот, где АЧХ исходного канала связи имеет глубокий провал.

Рассчитанный рассмотренным способом оптимальный фильтр минимизирует среднеквадратическую ошибку воспроизведения полезного сигнала. Эта ошибка состоит из двух статистически независимых составляющих:

1. Входного гауссова шума, прошедшего на выход фильтра.
2. Остаточной МСИ.

В результате оптимизации достигается оптимальный компромисс между подавлением этих двух составляющих благодаря минимизации их *суммарной дисперсии*. Поэтому для расчета эквалайзера, как видно из (3.12), необходимо знать отношение сигнал/шум. Используются также эквалайзеры, рассчитываемые без учета шума (при расчете принимается $\sigma_n^2 = 0$). Они минимизируют только остаточную МСИ и называются *эквалайзерами с форсированием нуля* (англ. термин — Zero Forcing Equalizer (ZFE)).

Кривые помехоустойчивости, полученные при приеме сигнала с модуляцией ФМ-2 с использованием линейных эквалайзеров (ZFE и рассчитанного с учетом уровня шума) 31-го порядка, в числе прочих вариантов показаны на приведенном ранее рис. 3.6 (штриховая и штрихпунктирная кривые соответственно). Видно, что проигрыш по сравнению со случаем отсутствия МСИ весьма велик. При этом ZFE-вариант существенно проигрывает эквалайзеру, рассчитанному с учетом уровня шума, и лишь при высоком отношении сигнал/шум кривые помехоустойчивости для этих двух эквалайзеров начинают сближаться (дисперсия шума при этом мала, и ее учет слабо влияет на результаты расчета фильтра).

Линейный эквалайзер дает плохие результаты для каналов, АЧХ которых имеет глубокие провалы (именно этим объясняется большой проигрыш, полученный в рассмотренном примере). Это связано с тем, что для компенсации МСИ в таких условиях АЧХ фильтра на частотах провалов должна иметь высокие подъемы, а их наличие вызывает сильное увеличение уровня шума на соответствующих частотах и, как следствие, резкое падение общего отношения сигнал/шум на выходе фильтра. Расчет эквалайзера с учетом уровня шума дает, как уже отмечалось, компромиссное решение: высота пиков АЧХ фильтра при этом ограничивается, чтобы не допустить чрезмерного роста уровня шума. Но, несмотря на это, итоговая дисперсия ошибки фильтрации остается высокой. Улучшить результаты позволяет переход к нелинейным схемам, рассматриваемым далее.

3.5. Эквалайзер с обратной связью по решению

Линейный эквалайзер, в принципе, можно реализовать и в виде рекурсивного фильтра с бесконечной импульсной характеристикой. Однако это эквивалентно лишь увеличению порядка нерекурсивного фильтра, так что улучшение характеристик оказывается только количественным.

Качественного изменения ситуации можно добиться, если заметить, что сквозная импульсная характеристика системы «канал — эквалайзер» имеет вид, показанный на рис. 3.10 (см. также рис. 3.9, а).

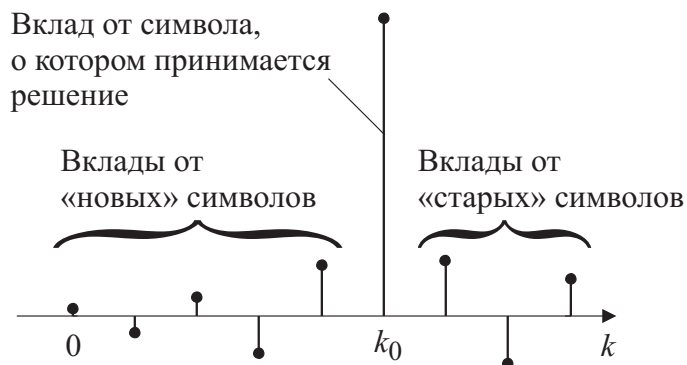


Рис. 3.10. Сквозная импульсная характеристика системы «канал — эквалайзер»

Пик сквозной импульсной характеристики, по уровню близкий к единице и расположенный в позиции k_0 , соответствует вкладу от символа, о котором демодулятор принимает решение. Часть ИХ слева от пика, — это вклады от символов, прошедших на выход с задержкой менее чем на k_0 тактов, т. е. от символов, решение о которых демодулятором еще не принято; часть ИХ справа от пика — это вклады от «старых» символов, решение о значении которых демодулятор уже принял. Но раз эти решения уже есть и известна сквозная импульсная характеристика, то можно попытаться скомпенсировать вклад в остаточную МСИ, обусловленный этими символами. Если отношение сигнал/шум достаточно велико, решения, принимаемые демодулятором, будут в основном правильными, и такая компенсация сможет существенно улучшить помехоустойчивость приема. Эта технология, широко применяемая на практике, называется обратной связью по решению (ОСР; англ. термин — Decision Feedback).

Итак, обратная связь по решению — это добавление к линейному эквалайзеру рекурсивной ветви, в которую подаются очищенные от шума оценки информационных символов, полученные в результате демодуляции (рис. 3.11). Эквалайзер, таким образом, состоит в данном случае из двух ветвей — прямой ветви с коэффициентами w_{FF} и ветви обратной связи с коэффициентами w_{FB} .

Эквалайзер с ОСР является *нелинейным* устройством, так как в цепь обратной связи входит нелинейный блок демодулятора, осуществляющий выбор ближайшей точки созвездия.

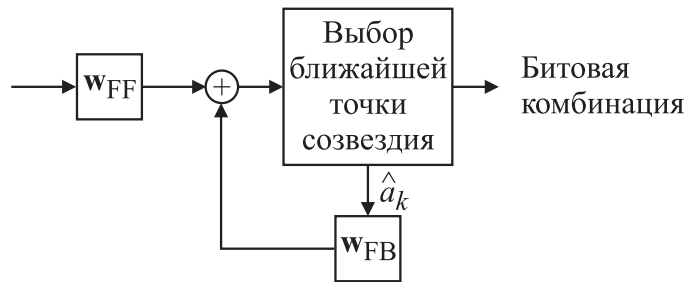


Рис. 3.11. Эквалайзер с обратной связью по решению

Если известна сквозная импульсная характеристика линейной части этой системы (канал — прямая ветвь эквалайзера), подобная показанной на рис. 3.10, то коэффициенты ветви обратной связи — это отсчеты сквозной ИХ, расположенные справа от ее пика, взятые с обратным знаком. Однако следует заметить, что прямая ветвь фильтра в данном случае отличается от используемой в линейном эквалайзере, так как при ее оптимизации не нужно принимать во внимание остаточную МСИ от предыдущих символов, скомпенсированную ветвью ОСР.

При расчете эквалайзера с ОСР обычно подразумевается, что в ветвь обратной связи всегда поступают правильные значения символов a_k , т. е. предполагается *идеальная* ОСР. При *реальной* ОСР решения, выдаваемые демодулятором, неизбежно содержат ошибки. Попадая в фильтр обратной связи, ошибочное значение влияет на результат фильтрации в течение нескольких символьных интервалов (их число определяется длиной фильтра обратной связи). В течение этого времени вероятность ошибок демодуляции будет повышенной — происходит *размножение ошибок*.

Многочисленные исследования показали, что проигрыш в помехоустойчивости при переходе от идеальной к реальной ОСР не превышает 2 дБ [2].

Кривые помехоустойчивости, полученные при приеме сигнала с модуляцией ФМ-2 с использованием эквалайзеров с идеальной и реальной ОСР, также показаны на приведенном ранее рис. 3.6 (сплошные кривые с маркерами). Графики демонстрируют существенное улучшение помехоустойчивости по сравнению с линейными эквалайзерами. Видна также разница примерно в 1 дБ между идеальной и реальной ОСР.

Несмотря на наличие данного эффекта, эквалайзеры с ОСР демонстрируют улучшение помехоустойчивости по сравнению с линейными эквалайзерами, особенно заметное в тех случаях, когда АЧХ канала содержит глубокие провалы.

3.6. Многочастотная модуляция (технологии OFDM, DMT)

Одной из наиболее популярных современных технологий передачи данных является ортогональное частотное мультиплексирование (Orthogonal Frequency Division Multiplexing, OFDM). Системы связи с OFDM позволяют

надежно и с высокой скоростью передавать данные по радиоканалу, в котором наблюдаются многолучевое распространение и замирания.

При OFDM общая полоса частот канала делится на некоторое количество узкополосных каналов, а исходный поток данных также делится на потоки с меньшей скоростью, передаваемые в отдельных каналах на разных несущих частотах (их называют *поднесущими* — subcarrier). Использование частотно-временного ресурса при применении одной и нескольких несущих частот иллюстрирует рис. 3.12. Из рисунка видно, что общее число символов, которое можно передать за время Δt в полосе W , равно $N = W\Delta t$ и не зависит от способа использования частотно-временного ресурса.

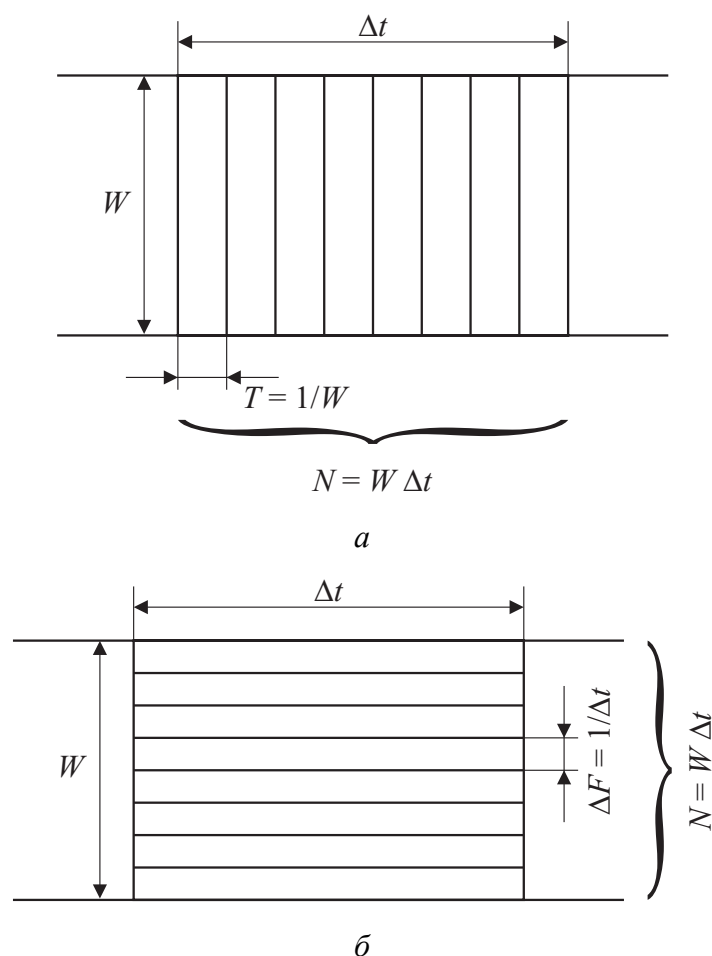


Рис. 3.12. Варианты использования частотно-временного ресурса:
 а — с одной несущей, б — с несколькими несущими

Уменьшение скорости передачи в отдельном канале позволяет увеличить длительность символа и, следовательно, ослабить влияние межсимвольной интерференции (рис. 3.13). На рис. 3.13, б показан сигнал с одной несущей и короткими символами, а на рис. 3.13, а — более длинный символ многочастотной системы. Длительность ИХ канала связи в обоих случаях одинакова, так что символы одинаково «размазываются» во времени, но вследст-

вие большей длительности символов в многочастотной системе их наложение друг на друга оказывается незначительным.

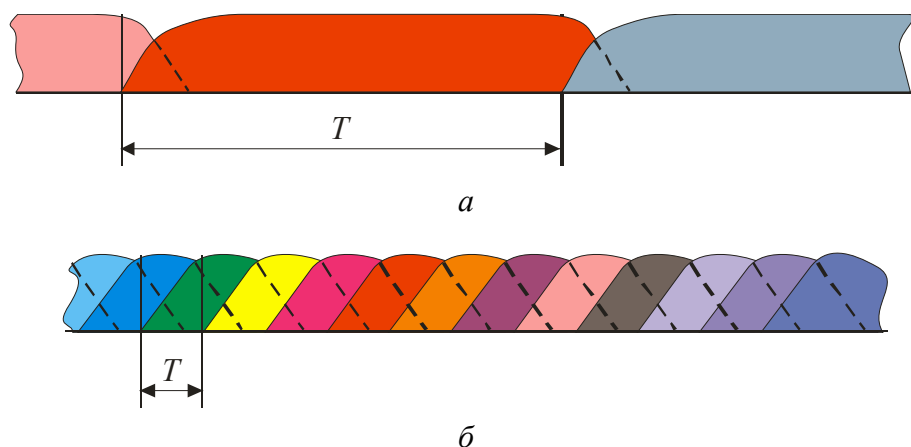


Рис. 3.13. Борьба с МСИ за счет многочастотной модуляции — иллюстрация во временной области

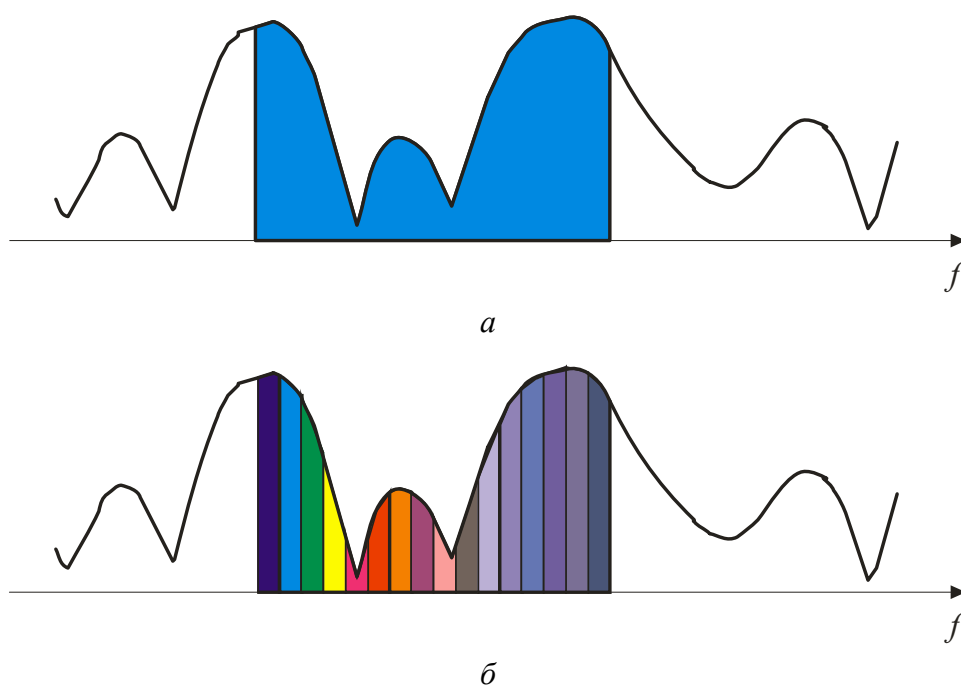


Рис. 3.14. Борьба с МСИ за счет многочастотной модуляции — иллюстрация в частотной области

Можно проиллюстрировать ослабление МСИ и в частотной области. Поскольку полоса частот каждого отдельного подканала узкая, влияние частотной дисперсии оказывается пренебрежимо мало и такой канал можно считать каналом с аддитивным белым гауссовским шумом (рис. 3.14). На рис. 3.14, *а* изображен спектр сигнала с одной несущей (он сильно искажен, так как АЧХ канала в пределах спектра сигнала сильно меняется). Рисунок

3.14, б иллюстрирует, что узкие спектры сигналов на отдельных поднесущих при той же АЧХ канала оказываются искажены лишь незначительно.

Примерами коммерческих систем, работающих с OFDM, являются беспроводные сети WiFi и WiMax (стандарты IEEE 802.11a/g и 802.16) и цифровое эфирное телевидение DVB-T (Digital Video Broadcasting — Terrestrial), а также его разновидность для приема на мобильные устройства DVB-H (Digital Video Broadcasting — Handheld). Также некоторые проводные системы xDSL (ADSL, VDSL) работают с одной из разновидностей OFDM, известной под названием DMT (Discrete MultiTone), без переноса на радиочастоту и с возможностью адаптивного распределения битового потока по поднесущим в зависимости от отношения сигнал/шум.

Принципиальное отличие OFDM от обычной многоканальной передачи с частотным разделением каналов заключается в том, что спектры сигналов, соответствующих отдельным узкополосным каналам, не разделены защитными частотными интервалами; напротив, они намеренно делаются перекрывающимися (расстояние между соседними поднесущими ΔF равно величине, обратной по отношению к длительности символа T : $\Delta F = 1/T$). Однако благодаря такому частотному расстоянию между соседними поднесущими сигналы разных каналов оказываются ортогональными (см. 2.2.6) и, следовательно, не создают взаимных помех при приеме.

СПМ сигнала на каждой поднесущей имеет форму $(\sin x/x)^2$, так как сигнальная посылка в данном случае является прямоугольной. В результате сложения таких спектров общий спектр OFDM-сигнала приближается по форме к прямоугольнику (рис. 3.15).

Для модуляции поднесущих в общем случае используется фазовая или квадратурная модуляция. При этом число бит, передаваемых в одном временном такте, может быть одинаковым или разным для разных поднесущих. В радиолиниях, как правило, используются одинаковые параметры модуляции для всех поднесущих, а в проводных каналах связи, где характеристики канала существенно зависят от частоты и почти не меняются во времени, широко используется адаптивное распределение информационной нагрузки по поднесущим (системы ADSL, VDSL).

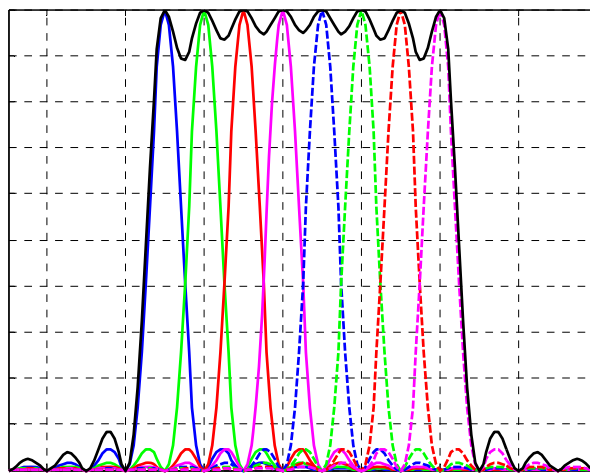


Рис. 3.15. Формирование спектра OFDM-сигнала

При использовании одинаковой модуляции на всех поднесущих помехоустойчивость в АБГШ-канале определяется видом модуляции и не зависит от числа поднесущих. Некоторые потери вызываются лишь добавлением циклического префикса и пилот-сигналов (см. далее).

Указанное ранее расстояние между поднесущими ($\Delta F = 1/T$) делает возможной вычислительно эффективную реализацию формирователя сигнала на основе быстрого преобразования Фурье (БПФ). Для пояснения этой идеи рассмотрим структуру OFDM-сигнала подробнее. Комплексную огибающую OFDM-сигнала можно представить в виде суммы сигналов нескольких ортогональных поднесущих с частотами ω_k (суммирование производится по номерам используемых поднесущих):

$$\dot{S}(t) = \sum_{(k)} \dot{A}_k(t) e^{j\omega_k t}. \quad (3.17)$$

Частоты поднесущих определяются исходя из расстояния между ними:

$$\omega_k = 2\pi k/T.$$

Комплексная огибающая $\dot{A}_k(t)$ каждой поднесущей представляет собой КАМ- либо ФМ-сигнал с прямоугольными сигнальными посылками. Рассмотрим поведение комплексной огибающей OFDM-сигнала (3.17) на длине одного символа (при этом $\dot{A}_k(t) = \dot{A}_k = \text{const}$):

$$\dot{S}_m(t) = \sum_{(k)} \dot{A}_k e^{j\omega_k t} = \sum_{(k)} \dot{A}_k e^{j \frac{2\pi k t}{T}}, \quad 0 \leq t < T.$$

Теперь дискретизируем этот сигнал по времени, выбрав период дискретизации так, чтобы на длительности символа T укладывалось целое число отсчетов. Обозначив это число как N , получим набор моментов времени $t_n = nT/N$ и отсчеты сигнала, равные

$$\dot{S}_m(t_n) = \sum_{(k)} \dot{A}_k e^{j \frac{2\pi k t_n}{T}} = \sum_{(k)} \dot{A}_k e^{j \frac{2\pi k n}{N}}, \quad 0 \leq n < N. \quad (3.18)$$

Видно, что получившийся символ (3.18) есть не что иное, как обратное дискретное преобразование Фурье от набора комплексных амплитуд \dot{A}_k (с точностью до постоянного множителя). Это и дает возможность реализовывать формулу (3.18) на практике с помощью экономных алгоритмов БПФ. Прием сигнала также может осуществляться с использованием БПФ.

При наличии межсимвольной интерференции между передаваемыми символами вводятся защитные интервалы, для заполнения которых используется *циклический префикс*: $\nu = \alpha N$ последних отсчетов добавляются перед началом символа (иными словами, набор отсчетов символа периодически

продолжается до заполнения защитного интервала, рис. 3.16). В результате, если сквозная импульсная характеристика канала связи будет короче, чем $\nu + 1$ отсчетов, *линейная* свертка передаваемого сигнала с импульсной характеристикой канала связи может быть представлена как свертка *циклическая* (начало символа будет искажено «наползающим» на него из-за наличия МСИ циклическим префиксом, т. е. «хвостом» этого же символа, рис. 3.17). Согласно свойствам дискретного преобразования Фурье [4], при этом Фурье-образ символа перемножается с Фурье-образом импульсной характеристики канала. Это позволяет эффективно реализовать компенсацию искажений в частотной области после БПФ, выполняемого в приемнике.

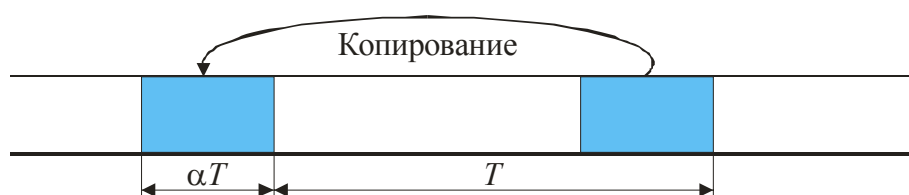


Рис. 3.16. Добавление циклического префикса



Рис. 3.17. Преобразование линейной свертки в циклическую

Добавление циклического префикса фактически означает увеличение длительности символов при сохранении значений всех прочих параметров сигнала. В результате ширина спектра сигнала на каждой поднесущей уменьшается, а суммарный спектр приобретает изрезанный характер. Такая гребенчатая форма спектра является характерной чертой OFDM-сигналов (рис. 3.18).

В реальных системах длина циклического префикса меняется в широких пределах — в зависимости от предполагаемых условий приема.

Так, в системе Wi-Fi длительность циклического префикса составляет 25 % от исходной длины OFDM-символа, а в стандарте цифрового радиовещания DRM эта длительность может меняться в диапазоне от $1/9$ до $11/14$.

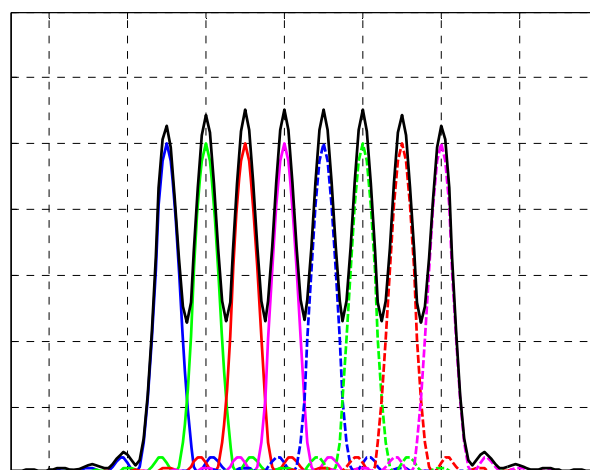


Рис. 3.18. Влияние циклического префикса на спектр OFDM-сигнала

Структурная схема передатчика OFDM-сигнала приведена на рис. 3.19. Прежде всего исходный битовый поток делится на подпотоки, предназначенные для передачи на отдельных поднесущих. Далее биты на каждой поднесущей группируются в символы нужного размера, и эти символы отображаются в точки используемых сигнальных созвездий. Обратное БПФ превращает набор комплексных амплитуд в частотной области в набор отсчетов OFDM-символа во временной области. После этого к сформированному символу добавляется циклический префикс и полученный набор отсчетов последовательно «вытаскивается» на цифроаналоговый преобразователь.

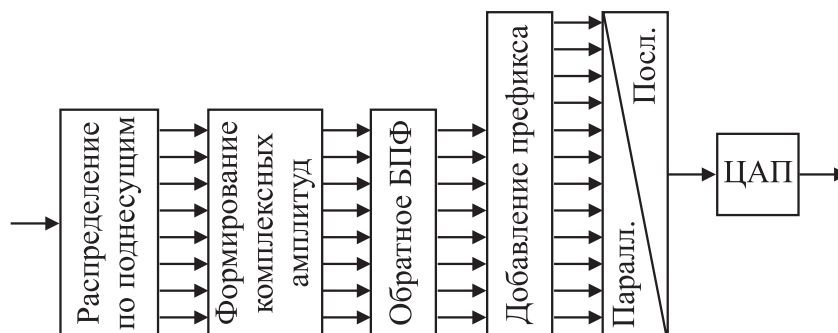


Рис. 3.19. Структура передатчика OFDM-сигнала

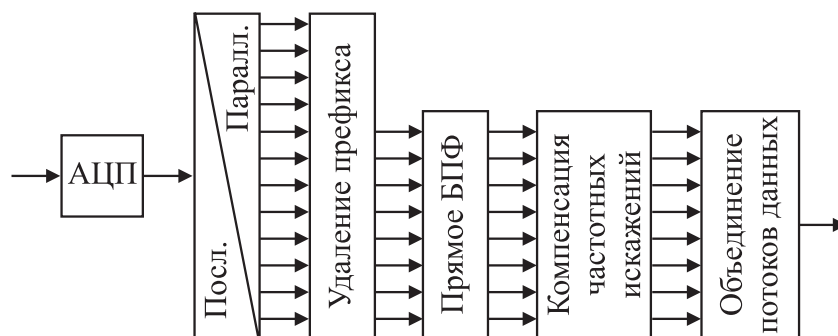


Рис. 3.20. Структура приемника OFDM-сигнала

Приемник OFDM-сигнала имеет зеркальную по отношению к передатчику структуру (рис. 3.20). Сигнал после аналого-цифрового преобразователя «нарезается» на символы, затем удаляется циклический префикс и вычисляется прямое БПФ. Полученные комплексные амплитуды разных поднесущих делятся на имеющиеся оценки комплексного коэффициента передачи канала связи на соответствующих частотах. Далее для скорректированных таким образом комплексных амплитуд ищутся ближайшие точки используемых сигнальных созвездий, затем демодулированные битовые комбинации объединяются, образуя итоговый принятый битовый поток.

Кратко рассмотрим количественные характеристики OFDM-сигнала. Пусть имеется N поднесущих, частотное расстояние между которыми составляет ΔF . Ширина спектра сигнала, таким образом, будет

$$W = N \Delta F.$$

Длительность одного символа, как уже отмечалось, составляет

$$T = 1/\Delta F.$$

Пусть защитный интервал составляет долю α от длительности символа. Тогда скорректированная длительность символов

$$T' = T(1 + \alpha).$$

Если на каждой поднесущей в одном символе передается m бит (предполагаем, что это число для всех поднесущих одинаково), то один OFDM-символ будет содержать Nm бит, а результирующая скорость передачи информации

$$R = \frac{Nm}{T'} = \frac{Nm}{T(1 + \alpha)} = \frac{N\Delta F m}{1 + \alpha} = \frac{W m}{1 + \alpha}. \quad (3.19)$$

Спектральная эффективность OFDM-системы

$$\frac{R}{W} = \frac{m}{1 + \alpha}. \quad (3.20)$$

Таким образом, как и в случае линейной модуляции с одной несущей, спектральная эффективность определяется числом бит в модуляционном символе (m), добавляются лишь потери из-за введения циклического префикса. Поэтому технология OFDM — это не средство искусственного расширения спектра сигнала, а способ использования частотного ресурса, позволяющий сравнительно просто реализовать в приемнике компенсацию эффектов многолучевого распространения радиоволн.

Для поддержания частотной и временной синхронизации, а также для оценивания характеристик канала связи в OFDM-сигнал вводятся *пилот-элементы* (рис. 3.21), содержимое которых генерируется по заранее заданному закону, известному на приемной стороне. В зависимости от их расположения на частотно-временной плоскости различают *пилот-тоны* (поднесущие, полностью отведенные для передачи пилот-элементов) и *пилот-символы* (OFDM-символы, полностью отведенные для передачи пилот-элементов). Кроме того, в ряде систем применяются *пилот-ячейки*, расположение которых не является фиксированным ни по времени, ни по частоте. Как правило, такие ячейки образуют на частотно-временной плоскости регулярную сетку (рис. 3.21). Это дополнительно снижает полезную скорость передачи информации и спектральную эффективность по сравнению со значениями, даваемыми формулами (3.19) и (3.20). Во многих случаях это снижение является незначительным, однако при быстрых изменениях состояния канала связи требуется частая сетка пилот-ячеек, и тогда потери в скорости передачи могут стать заметными.

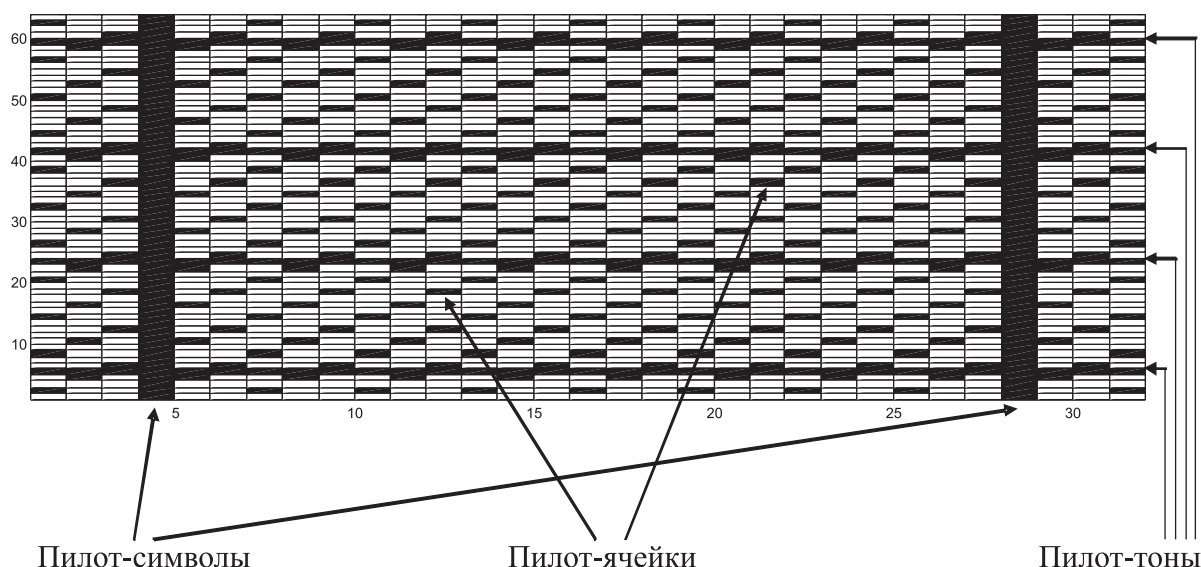


Рис. 3.21. Пилот-элементы в OFDM-сигнале

Достоинства и недостатки многочастотных систем. Достоинства технологий со многими несущими частотами многократно упоминались в тексте данного раздела. Кратко перечислим их еще раз:

- вычислительно эффективные методы формирования и приема сигнала;
- изящный способ борьбы с межсимвольной интерференцией.

Далее перейдем к обсуждению отрицательных сторон многочастотных систем.

Наиболее существенным недостатком OFDM-сигнала является большой пикфактор (отношение пиковой мощности к средней). Так как сигнал представляет сумму N независимых сигналов, соответствующих разным поднесущим, в силу центральной предельной теоремы распределение вероятности его значений приближается к нормальному. Распределение вероятности мгновенной мощности сигнала при этом оказывается экспоненциальным, из чего следует, что, например, превышение средней мощности в 2 раза (на 3 дБ) будет наблюдаться с вероятностью примерно 13,5 %. Столь существенные вариации амплитуды генерируемого сигнала требуют высокой линейности передающего тракта. В частности, в линейном режиме должен работать окончательный каскад передатчика, что снижает его КПД, не позволяя использовать энергетически эффективный режим работы с отсечкой.

Имеются многочисленные публикации о способах уменьшения пикфактора OFDM, как правило, за счет введения избыточности в частотной области — часть поднесущих используется исключительно для снижения пикфактора выбором соответствующих амплитуд и фаз колебаний. Однако значительно уменьшить пикфактор можно только за счет существенного снижения скорости передачи информации.

Кроме того, несмотря на использование вычислительно эффективных алгоритмов БПФ, устройства формирования и приема OFDM-сигнала требуют большой интенсивности вычислений, что ведет к сравнительно высокому энергопотреблению приемопередающих устройств. По этой причине до недавнего времени использование OFDM в мобильной аппаратуре связи было затруднено.

Итак, кратко список недостатков OFDM можно сформулировать следующим образом:

- высокое энергопотребление;
- большой пикфактор сигнала;
- необходимость использования усилителя мощности в линейном режиме.

4. ЗАМИРАНИЯ И РАЗНЕСЕНИЕ

Отличительной чертой большинства используемых на практике радиоканалов являются *замирания* — изменения условий распространения радиоволн во времени (исключением можно считать разве что космическую связь и радиорелейные линии с узконаправленными антеннами). В данном разделе будут рассмотрены математическое описание замираний и их влияние на помехоустойчивость приема.

Но, прежде чем перейти к обсуждению собственно замираний, рассмотрим основные энергетические соотношения, справедливые для радиолиний.

4.1. Энергетические соотношения в радиолинии

Пусть P_0 — излучаемая передатчиком мощность. При всенаправленной антенне мощность равномерно распределена по всем направлениям, поэтому на расстоянии d эта мощность равномерно распределяется по поверхности сферы соответствующего радиуса. Площадь поверхности сферы равна $4\pi d^2$, в итоге плотность потока мощности

$$p = \frac{P_0}{4\pi d^2}.$$

Диаграмма направленности антенны неравномерно распределяет поток мощности по разным направлениям. Величина, показывающая, во сколько раз поток мощности в нужном направлении оказывается больше, чем при всенаправленном излучении, называется *коэффициентом направленного действия* (КНД) антенны $G_{\text{ТХ}}$. С учетом этого поток мощности изменяется:

$$p = \frac{P_0 G_{\text{ТХ}}}{4\pi d^2}.$$

Мощность, принятая антенной, определяется произведением потока мощности на *эффективную площадь* антенны A_{RX} :

$$P_{\text{RX}} = p A_{\text{RX}} = \frac{P_0 G_{\text{ТХ}} A_{\text{RX}}}{4\pi d^2}.$$

Эффективная площадь A и КНД G — два связанных друг с другом параметра. При $A \gg \lambda^2$ ($\lambda = c/f$ — длина волны, f — частота, $c = 3 \cdot 10^8$ м/с — скорость света) эта связь выражается следующим образом:

$$G = \frac{4\pi A}{\lambda^2} = \frac{4\pi A f^2}{c^2}.$$

Таким образом, в зависимости от того, какие характеристики антенн используются, принятая мощность может быть выражена одним из следующих четырех способов:

1. $P_{\text{RX}} = P_0 \frac{G_{\text{ТХ}} A_{\text{RX}}}{4\pi d^2}.$

2. $P_{RX} = P_0 \frac{A_{TX} G_{RX}}{4\pi d^2}.$
3. $P_{RX} = P_0 \frac{A_{TX} A_{RX}}{\lambda^2 d^2} = P_0 \frac{A_{TX} A_{RX} f^2}{c^2 d^2}.$
4. $P_{RX} = P_0 \frac{G_{TX} G_{RX} \lambda^2}{(4\pi d)^2} = P_0 \frac{G_{TX} G_{RX} c^2}{(4\pi df)^2}.$

Следует обратить внимание на то, что эти формулы дают разную зависимость затухания мощности от частоты сигнала: в случаях 1 и 2 эта зависимость отсутствует, в случае 3 с ростом частоты затухание уменьшается (принятая мощность растет), а в случае 4 — увеличивается (принятая мощность падает). Для понимания этих различий полезно представить себе, что должно происходить с передающей и приемной антеннами, чтобы описывающие их параметры оставались постоянными при изменении частоты.

В случае 3 антенны характеризуются эффективной площадью, поэтому при увеличении частоты подразумевается постоянство этого параметра. Но эффективная площадь пропорциональна реальной площади антенны, поэтому будут оставаться постоянными и ее геометрические размеры. С ростом частоты отношение размера антенны к длине волны будет возрастать, в результате будет сужаться диаграмма направленности антенны. Передающая антенна будет все сильнее концентрировать мощность излучения в заданном направлении. Способность приемной антенны перехватывать мощность сигнала, приходящего с заданного направления, определяется эквивалентной площадью и поэтому не меняется. В результате мощность сигнала на входе приемника с ростом частоты будет повышаться.

В случае 4 постоянными должны являться коэффициенты направленного действия антенн, а это означает, что размеры антенн меняются прямо пропорционально длине волны (обратно пропорционально частоте). Эффективность передающей антенны в данном случае при изменении частоты не меняется, а вот у приемной антенны с ростом частоты (с уменьшением длины волны) будет уменьшаться эффективная площадь, что приведет к уменьшению мощности сигнала на входе приемника.

В случаях 1 и 2 при изменении частоты меняется размер только одной из антенн, что и приводит к отсутствию частотной зависимости мощности на входе приемника.

Безразмерный коэффициент из формулы, соответствующей варианту 4,

$$\left(\frac{4\pi d}{\lambda} \right)^2 = \left(\frac{4\pi df}{c} \right)^2,$$

называется *потерями в свободном пространстве*.

Спектральная плотность мощности теплового шума

$$N_0 = k T K_{\text{ш}},$$

где $k = 1,38 \cdot 10^{-23}$ Дж/К — постоянная Больцмана; T — абсолютная температура; $K_{\text{ш}}$ — коэффициент шума приемника.

Энергия принятого бита определяется принятой мощностью и скоростью передачи информации: $E_b = P_{\text{RX}}/R$. Отсюда получаем окончательную формулу для отношения сигнал/шум на бит:

$$\frac{E_b}{N_0} = P_0 \frac{G_{\text{TX}} G_{\text{RX}}}{(4\pi d/\lambda)^2 R k T K_{\text{ш}}}.$$

Замечание. Значения мощности в разных точках системы связи могут различаться на много порядков. Так, мощность, излучаемая передатчиком, может измеряться киловаттами, а мощность сигнала, поступающего на вход приемника, — пиковаттами. Поэтому во многих случаях оказывается удобно пользоваться логарифмическими единицами мощности — децибелами по отношению к некоторому опорному уровню. В качестве таких уровней используются 1 Вт (соответствующая логарифмическая единица — «децибел по отношению к ватту», дБВт, dBW) и 1 мВт («децибел по отношению к милливатту», дБм, дБмВт, dBm). Преобразование осуществляется по следующим формулам:

$$P_{[\text{дБВт}]} = 10 \lg \frac{P}{1 \text{ Вт}}, \quad P_{[\text{дБм}]} = 10 \lg \frac{P}{1 \text{ мВт}}.$$

Табл. 4.1. Связь между линейными и логарифмическими единицами мощности

Лин. ед.	1 кВт	2 Вт	1 Вт	0,5 Вт	0,1 Вт	10 мВт	1 мВт	1 мкВт	1 нВт	1 пВт
дБВт	30	3	0	−3	−10	−20	−30	−60	−90	−120
дБм	60	33	30	27	20	10	0	−30	−60	−90

В табл. 4.1 приведены примеры перевода мощности из линейных единиц в логарифмические.

4.2. Причины возникновения и общие характеристики замираний

Из-за отражений радиоволн от различных объектов сигнал попадает на приемную антенну несколькими разными путями (*лучами*). В результате сложения этих лучей формируется общий коэффициент передачи канала. При изменении условий распространения коэффициент передачи канала меняется. Этот процесс и называется *замираниями* (fading). Таким образом, основная причина возникновения замираний — многолучевое распространение радиоволн.

Различают *быстрые* и *медленные* замирания. Быстрые замирания связаны с изменением фазовых соотношений в принимаемых лучах. Для этого достаточно небольших (не превышающих длины волны) перемещений передатчика, приемника или отражающих объектов. Медленные замирания про-

исходят из-за изменения амплитуд лучей и/или их набора. Это связано со значительными относительными перемещениями объектов.

В зависимости от разброса времени прихода лучей $\Delta\tau$ выделяют также *селективные* и *неселективные* замирания.

При неселективных замираниях разброс времени прихода лучей $\Delta\tau \ll 1/W$, где W — ширина спектра сигнала. Лучи *неразрешимы* для приемника, они суммируются и воспринимаются им как один суммарный луч со случайными амплитудой и фазой.

В случае селективных замираний $\Delta\tau \gg 1/W$. Лучи *разрешимы*, приемник воспринимает их раздельно, это можно использовать при обработке сигнала (см. разд. 5 о расширении спектра).

Типичные значения разброса задержек зависят от условий распространения радиоволн:

- в помещении — 0,01...1 мкс (разность длин хода лучей — 3...300 м);
- в пригородах — 0,2...2 мкс (60...600 м);
- в городской застройке — 1...30 мкс (300 м...9 км).

Из-за замираний коэффициент передачи канала связи представляет собой случайную функцию времени и частоты. Более подробно математическая модель такого канала будет обсуждаться далее, а пока что отметим, что корреляционные свойства этого случайного процесса по временной и частотной осям характеризуются соответственно *временем когерентности* и *полосой когерентности* канала. Это, соответственно, интервал времени и частотный интервал, на протяжении которых значения коэффициента передачи имеют сильную статистическую связь.

Характеристики канала меняются во времени, что приводит к *мультипликативным искажениям* сигнала. Следовательно, происходят доплеровские сдвиги частоты (расширение спектральной линии). Значение максимального доплеровского сдвига обратно пропорционально времени когерентности канала и может быть рассчитано как

$$\Delta f_D = \frac{vf_0}{c}, \quad (4.1)$$

где v — скорость движения объекта; c — скорость света; f_0 — несущая частота системы связи.

4.3. Математическая модель канала с замираниями

При наличии многолучевого распространения радиоволн сигнал попадает от передатчика к приемнику по нескольким лучам, каждый из которых имеет свою задержку τ_k и комплексный коэффициент передачи \dot{h}_k . Из-за замираний эти величины меняются во времени, поэтому связь между переданным $s(t)$ и принятым $r(t)$ сигналами можно записать следующим образом:

$$r(t) = \sum_k \dot{h}_k(t) s(t - \tau_k(t)).$$

Функции $\dot{h}_k(t)$ и $\tau_k(t)$ представляют собой случайные процессы. На практике для упрощения анализа и моделирования набор задержек, как правило, считается постоянным, меняются во времени лишь комплексные коэффициенты передачи лучей.

В стандартах на системы связи обычно приводятся параметры моделей каналов связи для тестирования оборудования. Эти модели включают в себя задержки лучей, их относительные мощности, вероятностные (см. 4.4) и спектральные (см. 4.5) характеристики соответствующих случайных коэффициентов передачи. В табл. 4.2 в качестве примера приведены параметры одной из таких моделей, используемой для систем сотовой связи GSM (она соответствует распространению радиоволн в условиях городской застройки).

Табл. 4.2. Параметры модели многолучевого канала с замираниями

Номер луча	Задержка, мкс	Средняя относительная мощность, дБ	Распределение вероятности	Доплеровский спектр
1	0,0	−3,0	Рэлеевское	Классический (Джейкса)
2	0,2	0,0		
3	0,5	−2,0		
4	1,6	−6,0		
5	2,3	−8,0		
6	5,0	−10,0		

Для математического описания случайных коэффициентов передачи $\dot{h}_k(t)$ используется обычный математический аппарат анализа случайных процессов — вероятностные, корреляционные и спектральные характеристики. Рассмотрим их подробнее.

4.4. Вероятностные свойства коэффициента передачи канала связи

Рассмотрим распространенные статистические модели для комплексного коэффициента передачи \dot{h} канала с неселективными замираниями (они же могут использоваться для отдельных разрешимых лучей в случае селективных замираний).

1. Рэлеевский канал. В данном случае предполагается, что в точке приема суммируется много лучей со статистически независимыми параметрами и примерно одинаковым уровнем. Согласно центральной предельной теореме, при большом числе лучей их сумма будет иметь гауссово распределение вероятности.

Итак, \dot{h} в данном случае представляет собой комплексную нормальную случайную величину с нулевым средним значением. Плотность вероятности

$$p(\dot{h}) = \frac{1}{2\pi\sigma_h^2} \exp\left(-\frac{|\dot{h}|^2}{2\sigma_h^2}\right),$$

где σ_h^2 — дисперсия вещественной и мнимой частей \dot{h} . Общая дисперсия комплексной величины \dot{h} равна $2\sigma_h^2$.

Плотность вероятности для модуля величины \dot{h} описывается законом Рэлея, что и определило название данного канала:

$$p(|\dot{h}|) = \frac{|\dot{h}|}{\sigma_h^2} \exp\left(-\frac{|\dot{h}|^2}{2\sigma_h^2}\right), \quad |\dot{h}| \geq 0.$$

График плотности вероятности для закона Рэлея показан на рис. 4.1.

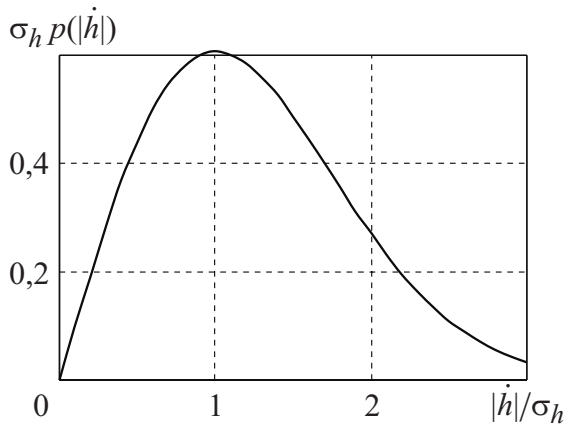


Рис. 4.1. Плотность вероятности модуля комплексного коэффициента передачи рэлеевского канала связи

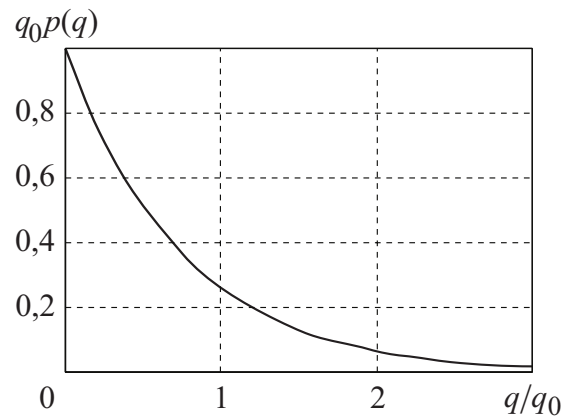


Рис. 4.2. Плотность вероятности мощности сигнала (или отношения сигнал/шум по мощности) в рэлеевском канале связи

Мощность принимаемого сигнала (а следовательно, и отношение сигнал/шум) имеет в этом случае экспоненциальное распределение:

$$p(q) = \frac{1}{q_0} \exp\left(-\frac{q}{q_0}\right), \quad q \geq 0, \quad (4.2)$$

где q_0 — математическое ожидание величины q . График этого распределения приведен на рис. 4.2. Формула (4.2) и рис. 4.2 показывают, что в рэлеевском канале наиболее вероятны значения мощности сигнала и отношения сигнал/шум, близкие к нулю.

2. Райсовский канал. Ситуация отличается от предыдущего случая тем, что имеется один луч, существенно превышающий остальные по своему уровню (обычно это луч *прямого распространения* — Line Of Sight, LOS). Ко всем остальным лучам по-прежнему можно применить центральную предельную теорему, в результате \dot{h} в данном случае также представляет собой комплексную нормальную случайную величину, но с *ненулевым* математическим ожиданием \dot{h}_0 , определяемым самым мощным лучом. Плотность вероятности

$$p(\dot{h}) = \frac{1}{2\pi\sigma_h^2} \exp\left(-\frac{|\dot{h} - \dot{h}_0|^2}{2\sigma_h^2}\right).$$

Плотность вероятности для модуля величины \dot{h} определяется законом Рэлея—Райса:

$$p(|\dot{h}|) = \frac{|\dot{h}|}{\sigma_h^2} \exp\left(-\frac{|\dot{h}|^2 + |\dot{h}_0|^2}{2\sigma_h^2}\right) I_0\left(\frac{|\dot{h}||\dot{h}_0|}{\sigma_h^2}\right), \quad |\dot{h}| \geq 0,$$

где $I_0(x)$ — модифицированная функция Бесселя первого рода нулевого порядка.

Можно ввести *коэффициент замираний* (fading parameter), характеризующий отношение мощности основного луча и средней мощности остальных лучей:

$$K = |\dot{h}_0|^2 / (2\sigma_h^2).$$

Графики плотности вероятности для закона Рэлея—Райса при нескольких значениях K показаны на рис. 4.3.

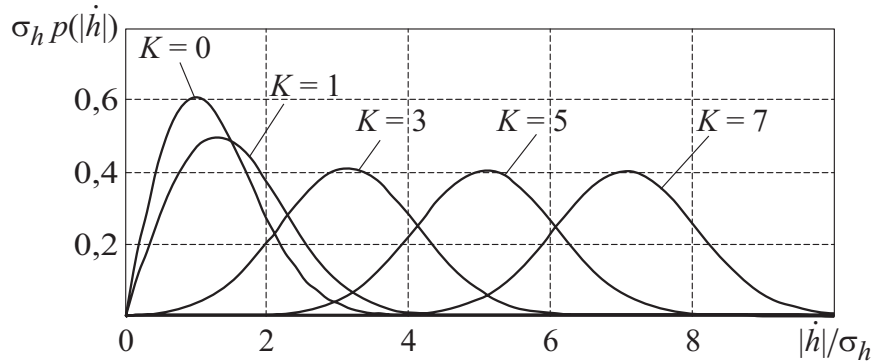


Рис. 4.3. Плотность вероятности модуля комплексного коэффициента передачи райсовского канала связи

При $K = 0$ основной луч отсутствует и получается рэлеевский канал. При $K \rightarrow \infty$ имеем только основной луч (замирания отсутствуют).

Средний коэффициент передачи по мощности в данном случае равен $\overline{|\dot{h}|^2} = |\dot{h}_0|^2 + 2\sigma_h^2$.

4.5. Корреляционные и спектральные характеристики коэффициента передачи канала связи

Динамика поведения случайного коэффициента передачи луча во времени (предполагаем, что этот случайный процесс является стационарным в широком смысле) описывается с помощью его корреляционной функции

$$R_h(\tau) = \dot{h}_k(t) \dot{h}_k^*(t - \tau) \quad (4.3)$$

и спектральной плотности мощности, согласно теореме Винера—Хинчина равной преобразованию Фурье от КФ:

$$H(f) = \int_{-\infty}^{\infty} R_h(\tau) e^{-j2\pi f \tau} d\tau. \quad (4.4)$$

СПМ (4.4) называется *доплеровским спектром*, так как эта функция описывает форму СПМ гармонического сигнала, искаженного данным каналом связи.

Эффективная протяженность КФ (4.3) по оси времени называется временем когерентности канала, а эффективная протяженность СПМ (4.4) по оси частот — *доплеровским разбросом частоты* (Doppler spread).

Наибольшее распространение при моделировании и теоретическом анализе каналов с замираниями получила *модель Джейкса* (Jakes), согласно которой КФ и СПМ описываются следующими формулами:

$$R_h(\tau) = \sigma_h^2 J_0(2\pi \Delta f_D \tau),$$

$$H(f) = \begin{cases} \frac{\sigma_h^2}{\pi \sqrt{\Delta f_D^2 - f^2}}, & |f| < \Delta f_D, \\ 0, & |f| \geq \Delta f_D, \end{cases}$$

где σ_h^2 — дисперсия коэффициента передачи; $J_0(x)$ — функция Бесселя первого рода нулевого порядка; Δf_D — доплеровский разброс, рассчитываемый по формуле (4.1) исходя из несущей частоты системы и предполагаемой максимальной скорости взаимного перемещения приемника и передатчика. Форма графиков КФ и СПМ показана на рис. 4.4.

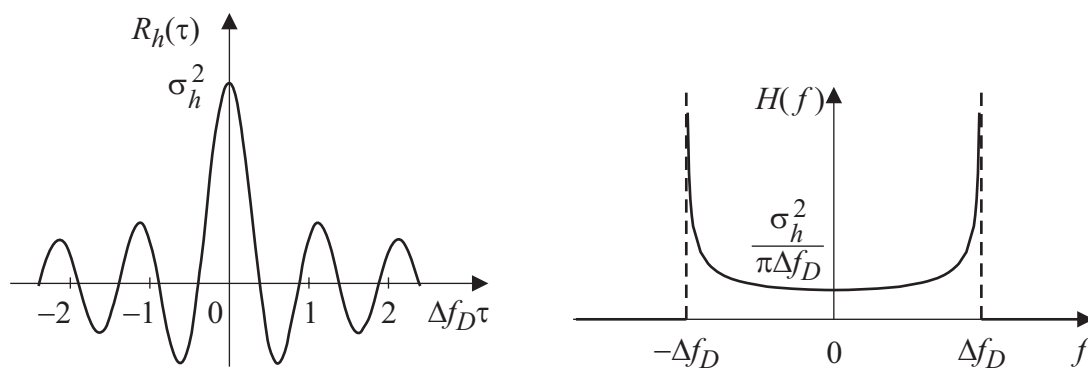


Рис. 4.4. Модель Джейкса: КФ (слева) и СПМ (справа)

Модель Джейкса распространена настолько широко, что в таблицах параметров моделей каналов, подобных табл. 4.2, этот спектр часто называют *классическим*.

4.6. Оценка помехоустойчивости различных видов модуляции в рэлеевском канале связи

Помехоустойчивость приема в условиях замираний зависит от ряда факторов. Самым главным из них является наличие информации о состоянии (коэффициенте передачи) канала (Channel State Information, CSI). Возможны следующие варианты:

- информация о канале отсутствует. При этом возможен только некогерентный прием частотной модуляции либо фазовой модуляции с дифференциальным кодированием;
- информация о канале имеется на приемной стороне (приемник может оценивать коэффициент передачи канала либо по тестовым фрагментам, предусмотренным в составе сигнала, либо «вслепую», зная только используемый вид модуляции);
- информация о канале имеется на приемной и передающей сторонах. Это возможно при наличии обратного канала связи и позволяет передатчику адаптивно регулировать параметры излучаемого сигнала. К сожалению, на практике состояние радиоканала обычно меняется слишком быстро, что делает данный подход нереализуемым.

Далее будут рассмотрены формулы, описывающие помехоустойчивость различных видов модуляции при наличии точной информации о состоянии канала связи на приемной стороне.

Для оценки помехоустойчивости в этих условиях необходимо усреднить вероятность ошибки, соответствующую АБГШ-каналу, с учетом статистических свойств отношения сигнал/шум в канале с замираниями:

$$P_b(q_0) = \int_0^{\infty} P_{\text{AWGN}}(q) p(q) dq, \quad (4.5)$$

где q — отношение сигнал/шум; $p(q)$ — его плотность вероятности; q_0 — математическое ожидание отношения сигнал/шум; $P_{\text{AWGN}}(q)$ — функция, описывающая вероятность ошибки для рассматриваемого вида модуляции в АБГШ-канале.

Аргументом, от которого зависит вероятность ошибки, в данном случае будет среднее отношение сигнал/шум q_0 .

Получим конкретные формулы помехоустойчивости нескольких видов модуляции в рэлеевском канале (т. е. для зависимости $p(q)$ будет использоваться формула (4.2)).

1. ФМ-2. Вероятность ошибки в АБГШ-канале, согласно (2.11), равна

$$P_b(q) = q \left(\sqrt{2q} \right). \quad (4.6)$$

Подставляя (4.6) и (4.2) в общее выражение (4.5), получаем

$$P_b(q_0) = \int_0^{\infty} Q(\sqrt{2q}) \frac{1}{q_0} e^{-q/q_0} dq = \frac{1}{2} \left(1 - \sqrt{\frac{q_0}{1+q_0}} \right). \quad (4.7)$$

Асимптотическое поведение этой зависимости при $q_0 \rightarrow \infty$ можно получить, воспользовавшись приближенным соотношением $\sqrt{1+x} \approx 1+x/2$, справедливым при $|x| \ll 1$:

$$P_b(q_0) = \frac{1}{2} \left(1 - \sqrt{1 - \frac{1}{1+q_0}} \right)_{q_0 \rightarrow \infty} \approx \frac{1}{2} \left(1 - \left(1 - \frac{1}{2+2q_0} \right) \right) = \frac{1}{4+4q_0} \approx \frac{1}{4q_0}. \quad (4.8)$$

График зависимости (4.7) показан на рис. 4.5 сплошной линией.

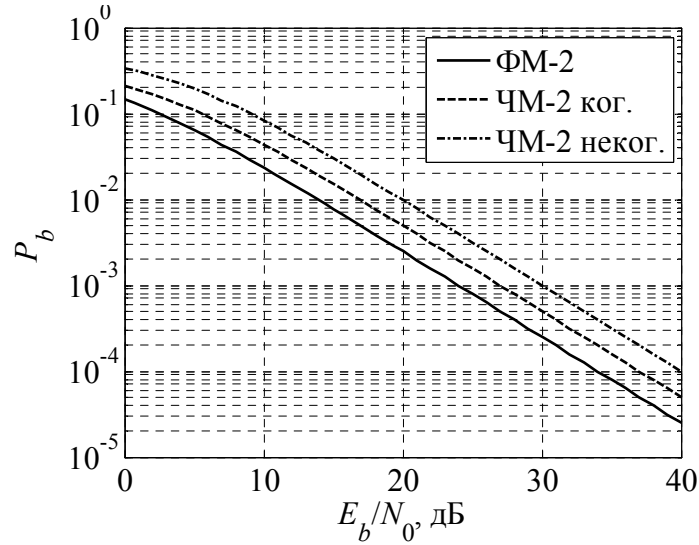


Рис. 4.5. Средняя помехоустойчивость в рэлеевском канале: ФМ-2 (сплошная линия), ортогональная ЧМ-2 с когерентным (штрихи) и некогерентным (штрихпунктир) приемом

2. ЧМ-2, когерентный прием ортогональных сигналов. Вероятность ошибки в АБГШ-канале, согласно (2.28), равна

$$P_b(q) = q(\sqrt{q}). \quad (4.9)$$

Подставляя (4.9) и (4.2) в общее выражение (4.5), получаем

$$P_b(q_0) = \int_0^{\infty} Q(\sqrt{q}) \frac{1}{q_0} e^{-q/q_0} dq = \frac{1}{2} \left(1 - \sqrt{\frac{q_0}{2+q_0}} \right). \quad (4.10)$$

При $q_0 \gg 1$ асимптотическое поведение (4.10) имеет вид

$$P_b(q_0) = \frac{1}{2} \left(1 - \sqrt{1 - \frac{2}{2+q_0}} \right)_{q_0 \rightarrow \infty} \approx \frac{1}{2} \left(1 - \left(1 - \frac{1}{2+q_0} \right) \right) = \frac{1}{4+2q_0} \approx \frac{1}{2q_0}. \quad (4.11)$$

График зависимости (4.10) показан на рис. 4.5 штриховой линией.

Сравнение формул (4.7) и (4.10), а также (4.8) и (4.11) показывает, что, как и в АБГШ-канале, переход от противоположных (ФМ-2) к ортогональ-

ным (ЧМ-2) сигналам приводит к двукратному энергетическому проигрышу (3 дБ).

3. ЧМ-2, некогерентный прием ортогональных сигналов. Вероятность ошибки в АБГШ-канале, согласно (2.30), равна

$$P_b(q) = \frac{1}{2} e^{-q/2}. \quad (4.12)$$

Подставляя (4.12) и (4.2) в общее выражение (4.5), получаем

$$P_b(q_0) = \int_0^{\infty} \frac{1}{2} e^{-q/2} \frac{1}{q_0} e^{-q/q_0} dq = \frac{1}{2 + q_0}. \quad (4.13)$$

Асимптотическое поведение этой формулы при $q_0 \gg 1$ очевидно:

$$P_b(q_0) = \frac{1}{2 + q_0} \approx \frac{1}{q_0} \quad q_0 \rightarrow \infty. \quad (4.14)$$

График зависимости (4.13) показан на рис. 4.5 штрихпунктирной линией.

Полученные формулы (4.8), (4.11) и (4.14) и приведенные на рис. 4.5 графики показывают катастрофическое влияние рэлеевских свойств канала на помехоустойчивость: вместо экспоненциального закона, справедливого для АБГШ-канала, вероятность ошибки убывает всего лишь пропорционально первой степени среднего отношения сигнал/шум!

4.7. Понятие вероятности отказа связи

Кроме средней помехоустойчивости, для характеристики работы системы связи при наличии замираний используется еще один параметр — *вероятность отказа связи* (outage probability). Под отказом связи подразумевается ситуация, когда качество приема падает ниже некоторого заданного уровня. Конкретное значение этого уровня определяется назначением системы связи и характером передаваемой информации.

Задав максимально допустимое значение вероятности ошибки $P_{b \max}$, можно определить (по формулам, относящимся к помехоустойчивости в АБГШ-канале) минимально допустимое отношение сигнал/шум q_{\min} , такое, что $P_{\text{AWGN}}(q_{\min}) = P_{b \max}$. Затем, зная плотность вероятности для отношения сигнал/шум в канале с замираниями $p_q(q)$, можно вычислить вероятность отказа связи по следующей формуле:

$$P_{\text{out}} = P(q < q_{\min}) = \int_0^{q_{\min}} p_q(q) dq.$$

4.8. Идея разнесения. Способы организации ветвей разнесенного приема

Улучшить качество приема сигнала в условиях замираний можно за счет организации *разнесенного приема* (англ. термин — *diversity*, дословно переводится как «разнообразие»). Идея разнесения состоит в реализации нескольких *ветвей* (branches) распространения сигнала от передатчика к приемнику таким образом, чтобы замирания сигнала в разных ветвях, а также аддитивные шумы разных ветвей были статистически независимы.

На качественном уровне причины выигрыша, достигаемого за счет разнесения, можно пояснить очень просто: так как замирания сигнала в разных ветвях независимы, вероятность того, что уровень сигнала сильно упадет во всех ветвях одновременно, мала.

В зависимости от способа реализации ветвей различают следующие варианты разнесения:

- *пространственное разнесение*. Это классический вариант разнесенного приема, при котором на приемной стороне используется несколько антенн, расположенных в разных точках пространства. Для независимости замираний в разных ветвях расстояние между ненаправленными приемными антеннами должно быть не меньше половины длины волны сигнала;
- *частотное разнесение*. Сигнал передается параллельно на нескольких несущих частотах. Для независимости замираний в разных ветвях разность частот должен превосходить полосу когерентности канала;
- *угловое разнесение*. За счет использования направленных приемных антенн можно разделять волны, приходящие из разных секторов пространства. Эти сигналы будут соответствовать различным путям распространения радиоволн от передатчика и за счет этого иметь независимые замирания;
- *временное разнесение*. Одна и та же информация передается повторно с задержкой по времени. Для независимости замираний в разных ветвях эта задержка должна превосходить время когерентности канала;
- *многолучевое разнесение*. Если отдельные лучи приема являются *разрешимыми* для приемника, т. е. могут восприниматься и обрабатываться отдельно, то они могут быть использованы в качестве ветвей разнесенного приема. Практическая реализация данной идеи будет обсуждаться в разд. 5, посвященном системам с расширением спектра.

4.9. Способы обработки принятых сигналов при разнесенном приеме

4.9.1. Общие формулы линейного комбинирования ветвей

Теперь рассмотрим, как можно объединять сигналы, поступающие по отдельным ветвям приема, чтобы получить выигрыш в помехоустойчивости.

Известно несколько способов такой обработки, различающихся сложностью практической реализации и достигаемыми характеристиками.

В качестве общей модели системы обработки будем использовать *линейное комбинирование* ветвей. Опишем математическую модель формирования итогового сигнала (конкретный способ реализации ветвей разнесения значения не имеет).

Пусть передатчиком излучается сигнал $s(t)$, имеется K ветвей разнесения и комплексный коэффициент передачи k -й ветви равен \dot{h}_k . Таким образом, полезный сигнал, принимаемый в k -й ветви, можно записать как

$$s_k(t) = s(t)\dot{h}_k, \quad k = 1, \dots, K.$$

СПМ белого шума в k -й ветви равна N_{0k} . В общем случае эти СПМ различны в разных ветвях приема, хотя далее часто будет предполагаться, что они одинаковы.

С учетом аддитивного шума $n_k(t)$ сигнал, принимаемый в k -й ветви, имеет вид

$$r_k(t) = s_k(t) + n_k(t) = s(t)\dot{h}_k + n_k(t), \quad k = 1, \dots, K.$$

Мерой отношения сигнал/шум в k -й ветви приема можно считать отношение

$$q_k = |\dot{h}_k|^2 / N_{0k}.$$

Напомним, что комплексный коэффициент передачи \dot{h}_k и, следовательно, отношение сигнал/шум q_k являются случайными величинами.

Приемная система вычисляет линейную комбинацию принимаемых сигналов, суммируя их с весовыми коэффициентами $\dot{\alpha}_k$:

$$s_\Sigma(t) = \sum_{k=1}^K \dot{\alpha}_k r_k(t) = s(t) \sum_{k=1}^K \dot{\alpha}_k \dot{h}_k + \sum_{k=1}^K \dot{\alpha}_k n_k(t).$$

Первая сумма в этом выражении дает итоговый уровень полезного сигнала. Так как шумы в разных ветвях являются независимыми, они суммируются по мощности. Таким образом, итоговое отношение сигнал/шум в выходном сигнале можно рассчитать следующим образом:

$$q_\Sigma = \frac{\left| \sum_{k=1}^K \dot{\alpha}_k \dot{h}_k \right|^2}{\sum_{k=1}^K N_{0k} |\dot{\alpha}_k|^2}. \quad (4.15)$$

Далее рассмотрим некоторые конкретные способы выбора коэффициентов $\dot{\alpha}_k$.

4.9.2. Выбор наиболее сильного луча

Выбор наиболее сильного луча — самый простой метод, при котором в каждый момент времени прием ведется только по одной ветви разнесения с номером k_0 — той, в которой отношение сигнал/шум максимально (т. е.

$$q_{k_0} = \max_k \{q_k\}):$$

$$\dot{\alpha}_k = \begin{cases} 1, & k = k_0, \\ 0, & k \neq k_0. \end{cases}$$

По-английски данный метод называется *Selection Combining* (SC). Его преимущество — простота реализации. Так как лучи в данном случае фактически не суммируются, нет нужды в их взаимной синхронизации и приемник является одноканальным.

Итоговое отношение сигнал/шум в данном случае равно максимуму отношений сигнал/шум в отдельных ветвях:

$$q_\Sigma = \max_k \{q_k\}.$$

Как известно, функция распределения для максимума из набора независимых случайных величин равна произведению их функций распределения, поэтому

$$F_{q_\Sigma}(q) = \prod_{k=1}^K F_{q_k}(q), \quad (4.16)$$

где $F_{q_k}(q)$ — функция распределения для отношения сигнал/шум в k -й ветви. Если статистические свойства отношений сигнал/шум во всех ветвях одинаковы, (4.16) принимает вид

$$F_{q_\Sigma}(q) = \left(F_q(q)\right)^K.$$

Дифференцирование этого выражения дает плотность вероятности итогового отношения сигнал/шум:

$$p_{q_\Sigma}(q) = \frac{dF_{q_\Sigma}(q)}{dq} = K \left(F_q(q)\right)^{K-1} p_q(q). \quad (4.17)$$

В случае рэлеевского канала $p_q(q)$ описывается формулой (4.2), а $F_q(q) = 1 - \exp(-q/q_0)$, $q \geq 0$. Подставив это выражение в (4.17), получаем

$$p_{q_\Sigma}(q) = \frac{K}{q_0} \left(1 - \exp\left(-\frac{q}{q_0}\right)\right)^{K-1} \exp\left(-\frac{q}{q_0}\right), \quad q \geq 0.$$

Расчет среднего значения этой величины дает

$$\overline{q_\Sigma} = \int_0^\infty q p_{q_\Sigma}(q) dq = \int_0^\infty q \frac{K}{q_0} \left(1 - \exp\left(-\frac{q}{q_0}\right)\right)^{K-1} \exp\left(-\frac{q}{q_0}\right) dq = q_0 \sum_{k=1}^K \frac{1}{k}.$$

Таким образом, добавление второго луча увеличивает среднее отношение сигнал/шум в $1 + 1/2 = 1,5$ раза, третьего — в $1 + 1/2 + 1/3 = 1,83$ раза и т. д. Каждый следующий луч увеличивает среднее отношение сигнал/шум, но все в меньшей степени.

Для расчета средней помехоустойчивости необходимо воспользоваться формулой (4.5). К сожалению, в большинстве случаев аналитический результат получить не удастся. Исключением является некогерентный прием ЧМ-2. В частности, при использовании двух и трех ветвей приема в этом случае получается следующее:

$$\overline{P_b} = \frac{4}{q_0^2 + 6q_0 + 8} \quad (K = 2), \quad (4.18)$$

$$\overline{P_b} = \frac{24}{q_0^3 + 12q_0^2 + 44q_0 + 48} \quad (K = 3). \quad (4.19)$$

Сравнивая эти результаты с формулой (4.13), видим, что с ростом среднего отношения сигнал/шум q_0 вероятность ошибки падает существенно быстрее — пропорционально степени величины q_0 , равной числу ветвей разнесения K .

4.9.3. Оптимальное сложение лучей

При оптимальном сложении лучей используются все ветви приема, а коэффициенты сложения $\dot{\alpha}_k$ выбираются таким образом, чтобы обеспечить максимальное отношение сигнал/шум (4.15) на выходе:

$$q_\Sigma = \frac{\left| \sum_{k=1}^K \dot{\alpha}_k \dot{r}_k \right|^2}{\sum_{k=1}^K N_{0k} |\dot{\alpha}_k|^2} \rightarrow \max. \quad (4.20)$$

По-английски данный вариант называется *Maximum Ratio Combining* (MRC).

Решение оптимизационной задачи (4.20) дает следующий результат (с точностью до произвольного постоянного множителя):

$$\dot{\alpha}_k = \frac{\dot{h}_k^*}{N_{0k}}. \quad (4.21)$$

При одинаковых уровнях шумов во всех ветвях разнесения результат упрощается:

$$\dot{\alpha}_k = \dot{h}_k^* . \quad (4.22)$$

Таким образом, при оптимальном сложении ветвей необходимо компенсировать фазовые сдвиги принимаемых сигналов, а модули коэффициентов сложения должны быть пропорциональны уровням полезных сигналов.

Полученные формулы (4.21) и (4.22) не случайно напоминают формулы для частотной характеристики согласованного фильтра [4, 5.1, 5.4] — согласованный фильтр максимизирует отношение сигнал/шум, оптимальным образом суммируя энергию различных участков спектра, что аналогично оптимальному суммированию сигналов разных ветвей разнесения.

Итоговое отношение сигнал/шум (4.15) после подстановки оптимальных коэффициентов (4.21):

$$q_{\Sigma} = \frac{\left| \sum_{k=1}^K \dot{\alpha}_k \dot{h}_k \right|^2}{\sum_{k=1}^K N_{0k} |\dot{\alpha}_k|^2} = \frac{\left| \sum_{k=1}^K \frac{\dot{h}_k^*}{N_{0k}} \dot{h}_k \right|^2}{\sum_{k=1}^K N_{0k} \left| \frac{\dot{h}_k^*}{N_{0k}} \right|^2} = \frac{\left| \sum_{k=1}^K q_k \right|^2}{\sum_{k=1}^K q_k} = \sum_{k=1}^K q_k .$$

Таким образом, при оптимальном сложении отношения сигнал/шум отдельных ветвей *суммируются*. При идентичных статистических свойствах замираний в разных ветвях среднее итоговое отношение сигнал/шум будет пропорционально числу ветвей:

$$\overline{q_{\Sigma}} = q_0 K .$$

Для получения плотности вероятности итогового отношения сигнал/шум нужно воспользоваться тем фактом, что ПВ суммы независимых случайных величин равна свертке их ПВ. Таким образом, например, в случае рэлеевского канала необходимо вычислить многократную свертку экспоненциальной ПВ (4.2) для его отношения сигнал/шум с самой собой. Результат имеет следующий вид:

$$p_{q_{\Sigma}}(q) = \frac{q^{K-1} \exp(-q/q_0)}{q_0^K (K-1)!}, \quad q \geq 0.$$

Для расчета средней помехоустойчивости необходимо воспользоваться формулой (4.5). В данном случае, в отличие от варианта с выбором максимально сильного луча, удастся получить много аналитических результатов. В частности, для модуляции ФМ-2 получается следующее:

$$\overline{P_b} = \int_0^{\infty} Q(\sqrt{2q}) p_{q_{\Sigma}}(q) dq = \left(\frac{1-\gamma}{2} \right)^K \sum_{k=0}^{K-1} C_{K-1+k}^k \left(\frac{1+\gamma}{2} \right)^k, \quad (4.23)$$

где $\gamma = \sqrt{q_0/(1+q_0)}$.

В случае когерентного приема ЧМ-2 результат описывается той же формулой (4.23), но при этом $\gamma = \sqrt{q_0/(2 + q_0)}$.

В случае некогерентного приема ЧМ-2 результат оказывается следующим:

$$\overline{P_b} = \int_0^\infty \frac{1}{2} \exp\left(-\frac{q}{2}\right) p_{q_\Sigma}(q) dq = \frac{1}{2\left(1 + \frac{q_0}{2}\right)^K}. \quad (4.24)$$

Асимптотически, при $q_0 \gg 1$, в обеих формулах (4.23) и (4.24) вероятность ошибки падает пропорционально K -й степени величины q_0 (это наблюдалось и при выборе наиболее сильного луча, см. (4.18) и (4.19)). Это означает, что при использовании рассмотренных способов объединения ветвей реализуется *полный порядок разнесения*.

4.9.4. Сложение с равными весами

Сложение с равными весами — упрощенный вариант сложения, при котором только компенсируются фазовые сдвиги, а модули коэффициентов суммирования не зависят от уровня сигналов в ветвях:

$$\dot{\alpha}_k = \exp\left(-j \arg(\dot{h}_k^*)\right).$$

По-английски данный вариант называется *Equal Gain Combining* (EGC). Техническая реализация данного метода проще, а показатели качества, естественно, несколько хуже, чем при оптимальном сложении. Типичный энергетический проигрыш по сравнению с оптимальным вариантом составляет около 1 дБ.

4.10. Перемежение как способ борьбы с группированием ошибок

При наличии замираний провалы коэффициента передачи занимают некоторое время, в течение которого вероятность ошибок оказывается выше средней. Это приводит к появлению *пакетных* ошибок, при которых искажается большое количество идущих подряд символов. Пакет ошибок может легко привести к тому, что корректирующая способность используемого кода окажется превышена. Помимо замираний в радиоканале к аналогичным эффектам приводят царапины на поверхности дисковых носителей информации.

Идея *перемежения* (interleaving) состоит в том, что перед поступлением в канал связи передаваемые символы определенным образом переставляются, а на приемной стороне их последовательность восстанавливается. Благодаря перемежению пакетная ошибка, произошедшая в канале связи, превращается в группу одиночных ошибок, распределенных на большом промежутке времени. В результате в каждый блок используемого кода попадает мень-

шее количество ошибок, и вероятность превышения корректирующей способности кода уменьшается.

Классификация перемежителей. Существует целый ряд алгоритмов перемежения, которые можно разделить на 2 больших класса — блочные и сверточные. При использовании блочных алгоритмов последовательность символов делится на блоки и перестановка символов производится в пределах каждого блока независимо. Сверточные алгоритмы предполагают непрерывную перестановку символов, без возможности выделить независимые блоки.

В любом случае использование пары «перемежитель—деперемежитель» вносит в поток данных некоторую задержку.

Блочные перемежители. Как уже отмечалось, в блочных перемежителях перестановка символов производится в пределах блока. Таким образом, в общем случае блочный перемежитель описывается *таблицей перестановки*, определяющей, на какое место в выходном блоке попадает каждый символ входного блока.

Распространенный вариант блочного перемежителя — *матричный* перемежитель. При использовании этого алгоритма блок представляется в виде прямоугольной матрицы, запись символов в эту матрицу производится последовательно вдоль строк, а считывание — вдоль столбцов, за счет чего и достигается эффект перестановки. Для деперемежения операции записи и считывания выполняются наоборот — запись производится вдоль столбцов, а считывание вдоль строк.

Сверточные перемежители. Термин «сверточный» используется по аналогии со сверточными кодами (см. разд. 8), математическая операция свертки к перемежителям отношения не имеет.

В общем случае сверточный перемежитель можно описать *таблицей задержек*, с которыми каждый символ входного потока данных попадает в выходной поток. Последовательность значений этих задержек является периодически повторяющейся. Деперемежение выполняется аналогичным образом, таблица задержек для этого выбирается так, чтобы компенсировать внесенный перемежителем разброс задержек, сделав задержки всех символов одинаковыми.

Распространенным вариантом является сверточный перемежитель с *линейным изменением задержки*. Такой алгоритм используется, в частности, в системах ADSL. При этом первый символ цикла повторения задержек попадает на выход без задержки, второй — с задержкой на Δk символов, третий — с задержкой на $2\Delta k$ символов и т. д.

5. РАСШИРЕНИЕ СПЕКТРА

5.1. Общие принципы расширения спектра

5.1.1. Понятие о расширении спектра

В разд. 2 было показано, что при использовании линейной модуляции минимально достижимая ширина спектра сигнала равна символьной скорости: $W_{\min} = R_{\text{sym}}$. Однако оказывается, что целый ряд положительных эффектов можно получить, если искусственно расширить полосу частот, занимаемую сигналом, преднамеренно снизив, таким образом, спектральную эффективность системы. Системы связи, использующие указанный подход, называются системами с *расширением спектра* (англ. термин — spread spectrum (SS)), что дословно переводится как «размазанный спектр»).

Однако простого сравнения ширины спектра сигнала со *средней* скоростью передачи символов оказывается недостаточно для отнесения системы к классу систем с расширением спектра. Дело в том, что расширение занимаемой сигналом полосы частот может быть достигнуто, в частности, за счет применения пакетной (прерывистой) передачи. Скорость передачи данных в пакете может во много раз превышать среднюю скорость, ширина спектра сигнала окажется соответствующим образом увеличена, но никаких специфических технологий расширения спектра при этом не будет использовано.

Подробное обсуждение проблем и неоднозначностей, связанных с определением понятия «расширение спектра», содержится, например, в [3, 1.1]. Здесь же, объединяя положения из [3, 1.1] и [1, 12.1], сформулируем следующее определение.

Системой с расширением спектра называется система связи, обладающая следующими свойствами:

- сигнал занимает *полосу частот*, существенно большую, чем необходимо для передачи информации с данной скоростью при использовании данного вида модуляции;
- расширение спектра производится с помощью некоторого *расширяющего кода*, независимого от передаваемых данных и представляющего собой *сложный сигнал* (сигнал с большой базой¹);
- в приемнике производится обратное сужение спектра сигнала с использованием *синхронизированной копии* расширяющего кода.

¹ *Базой* сигнала (Time-BandWidth product, TBW) называется произведение эффективных значений его длительности и ширины спектра: $B = \Delta f \Delta t$. Сложными сигналами (синонимы: широкополосные сигналы, сигналы с большой базой) называются сигналы, для которых $B \gg 1$ (см. [3], [4]).

5.1.2. Преимущества от использования расширения спектра

Краткий список преимуществ, которые может дать использование систем с расширением спектра, выглядит следующим образом:

- устойчивость к помехам;
- энергетическая скрытность;
- электромагнитная совместимость;
- структурная скрытность;
- новые возможности множественного доступа;
- устойчивость к замираниям и многолучевому распространению радиоволн;
- новые возможности разнесения.

Очертим в максимально общем виде механизм реализации указанных преимуществ. Далее в этом разделе способы реализации перечисленных преимуществ будут более подробно описаны применительно к конкретным технологиям расширения спектра.

Устойчивость к помехам. Чтобы описать механизмы подавления помех в системах с расширением спектра, не детализируя их структуру, воспользуемся подходом, изложенным в [3, 3.1]. Раздельно рассмотрим случаи узкополосной и широкополосной помех.

Узкополосная помеха. Пусть на систему связи действует помеха с мощностью J и полосой $W_J < W$. Оценим отношение сигнал/(шум плюс помеха) на выходе СФ приемника. При этом следует учесть, что у приемной стороны есть возможность адаптивно менять параметры приемника, чтобы минимизировать результаты воздействия помехи.

Сначала рассмотрим ситуацию, когда приемник не предпринимает никаких мер по борьбе с помехой. При отсутствии помехи отношение сигнал/шум по мощности на выходе СФ:

$$q^2 = \frac{2E}{N_0} = \frac{2EW}{N_0W}, \quad (5.1)$$

где E — энергия полезного сигнала. Если считать, что амплитудный спектр полезного сигнала близок к прямоугольной форме, такую же форму будет иметь и АЧХ СФ. Считая эту АЧХ равной единице в полосе пропускания фильтра, можно заметить, что знаменатель итоговой формулы (5.1) — это мощность шума на выходе фильтра. Также из предположения о характере АЧХ следует, что узкополосная помеха, расположенная в полосе сигнала, пройдет на выход СФ, не изменив своей мощности (рис. 5.1, а).

Помеха статистически не связана с полезным сигналом, поэтому снижение помехоустойчивости в данном случае будет вызвано только увеличением эквивалентной мощности шума на выходе СФ. В итоге отношение сигнал/(шум плюс помеха) на выходе СФ в данном случае

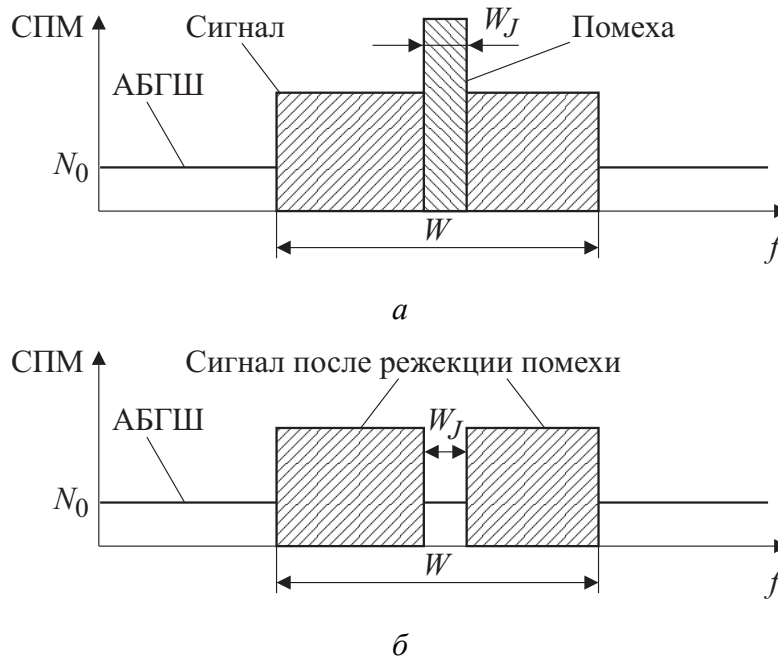


Рис. 5.1. Подавление узкополосной помехи в системе с расширением спектра

$$q_I^2 = \frac{2EW}{N_0W + J} = \frac{2E}{N_0 + J/W}. \quad (5.2)$$

Из этой формулы видно, что влияние помехи при прочих равных условиях будет тем меньше, чем больше полоса сигнала W , что и требовалось показать.

Теперь рассмотрим второй вариант, когда приемник реагирует на появление помехи, применяя режекторный фильтр, адаптивно настраиваемый на подавление полосы частот, занятой помехой (рис. 5.1, б). При идеальной режекции помеха не попадет на выход приемного фильтра и, следовательно, ее мощность не будет влиять на итоговое отношение сигнал/шум. Однако отношение сигнал/шум уменьшится из-за того, что в данном случае режекторный фильтр удалит и часть спектра полезного сигнала, вследствие чего его энергия уменьшится. Считая спектр сигнала прямоугольным, можно сказать, что снижение энергии полезного сигнала пропорционально доле режестируемой полосы частот. Таким образом, отношение сигнал/шум на выходе СФ

$$q_I^2 = \frac{2E}{N_0} \left(1 - \frac{W_J}{W} \right) = q^2 \left(1 - \frac{W_J}{W} \right).$$

Из формулы видно, что в данном случае увеличение ширины спектра W также приводит к ослаблению влияния помехи.

Широкополосная помеха. В отличие от узкополосных помех, которые могут быть непреднамеренно вызваны другими радиотехническими системами, широкополосные помехи чаще носят преднамеренный характер. При этом постановщик помех вынужден размазывать свою мощность по всей по-

лосе частот подавляемого сигнала, чтобы исключить возможность режекции пораженной полосы частот.

Отношение сигнал/(шум плюс помеха) в этом случае, как и при воздействии узкополосной помехи, определяется формулой (5.2), в которую входит СПМ помехи $N_J = J/W$. Так как энергетический ресурс (т. е. мощность J) постановщика помех ограничен, увеличение ширины спектра сигнала W приведет к снижению СПМ помехи N_J и, как следствие, к уменьшению ее влияния.

Энергетическая скрытность. За счет расширения спектра снижается спектральная эффективность системы, а вместе с ней и отношение сигнал/шум по мощности (см. формулу (1.2), связывающую отношение сигнал/шум на бит и отношение сигнал/шум по мощности). В результате сигнал «прячется» под шумами во временной области. Кроме того, расширение спектра приводит к снижению уровня СПМ сигнала (так как средняя мощность остается неизменной), что «прячет» сигнал под уровнем аддитивного белого шума в частотной области.

В результате сигнал может быть скрыт под шумами одновременно во временной и частотной областях, что затрудняет его обнаружение посторонним наблюдателем.

Электромагнитная совместимость. Можно назвать это преимущество «применением энергетической скрытности в мирных целях». В настоящее время частотно-временной ресурс используется многими радиотехническими системами одновременно. Чем сильнее расширен спектр сигнала, тем меньше уровень его СПМ и тем слабее, следовательно, оказываются помехи, создаваемые данной системой другим системам связи.

Структурная скрытность. Этот аспект важен для обеспечения информационной безопасности. Даже если сигнал обнаружен, для извлечения из него информации необходимо знание использованного расширяющего кода. Если параметры системы таковы, что возможно большое количество вариантов этого кода, «взлом» системы путем подбора кода может потребовать больших временных и материальных затрат.

Новые возможности множественного доступа. Множественному доступу к среде передачи и разделению каналов будет посвящен разд. 6, здесь же только заметим, что системы, реализующие расширение спектра, могут одновременно работать в одной и той же полосе частот с малыми взаимными помехами (или вообще без них) за счет использования разных расширяющих кодов. Это можно назвать «использованием структурной скрытности в мирных целях». Ансамбли ортогональных сигналов используются в этом случае не для M -ичной передачи, а для организации множественного доступа.

Устойчивость к замираниям и многолучевому распространению радиоволн. На рис. 5.2 показан пример возможной АЧХ многолучевого канала

связи и спектры двух сигналов: без расширения и с расширением спектра (рисунок демонстрирует только ширину спектров, уменьшение СПМ при расширении спектра на нем не показано).

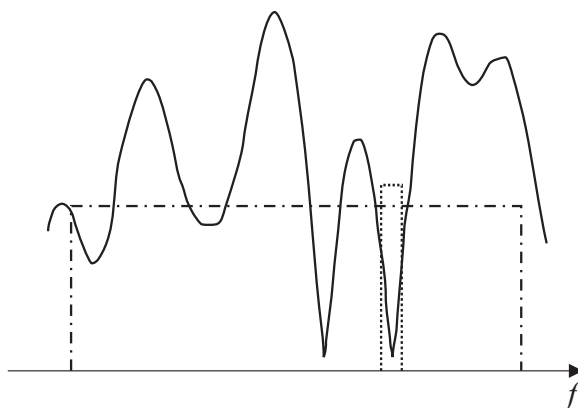


Рис. 5.2. АЧХ многолучевого канала (сплошная линия) и спектры сигнала без расширения спектра (пунктир) и с расширением спектра (штрихпунктир)

Из рисунка видно, что при одних и тех же параметрах канала связи вероятность сильного падения мощности принимаемого сигнала тем меньше, чем шире его спектр. Иными словами, если ширина спектра сигнала сильно превышает полосу когерентности канала связи, в пределах полосы сигнала АЧХ канала будет сильно варьироваться, демонстрируя как провалы, так и пики. В результате вероятность того, что коэффициент передачи канала окажется малым во всей полосе частот сигнала одновременно, существенно снижается. Таким образом системы с расширением спектра могут противостоять замираниям сигнала.

Что касается многолучевого распространения радиоволн, здесь главным является тот факт, что увеличение полосы сигнала может сделать отдельные лучи многолучевого канала *разрешимыми*, т. е. отдельно воспринимаемыми приемником (в разд. 4 было приведено условие разрешимости лучей: $\Delta\tau \gg 1/W$, где $\Delta\tau$ — разброс времени прихода лучей).

Новые возможности разнесения. Это преимущество является следствием устойчивости к многолучевому распространению: если отдельные лучи многолучевого канала являются разрешимыми, они могут по отдельности обрабатываться приемником и использоваться в качестве разных ветвей разнесения. Таким образом, в данном случае реализуется *многолучевое разнесение*.

5.1.3. Классификация систем с расширением спектра

На практике чаще всего используются два способа расширения спектра:

- прямое расширение спектра (англ. термин — Direct Sequence Spread Spectrum (DSSS));

- псевдослучайная перестройка рабочей частоты (ППРЧ; англ. термин — Frequency Hopping Spread Spectrum (FHSS)).

Далее рассматривается сущность этих технологий и механизмы реализации перечисленных ранее преимуществ при их использовании.

5.2. Прямое расширение спектра

5.2.1. Принципы реализации прямого расширения спектра

Сущность прямого расширения спектра состоит в том, что модулированный сигнал $d(t)$ в передатчике умножается на расширяющую последовательность $p(t)$, удовлетворяющую условию $|p(t)| = 1$ (рис. 5.3):

$$s(t) = d(t) p(t).$$

В принципе, вид модуляции может быть любой, но на практике прямое расширение спектра обычно используется в сочетании с линейной модуляцией.

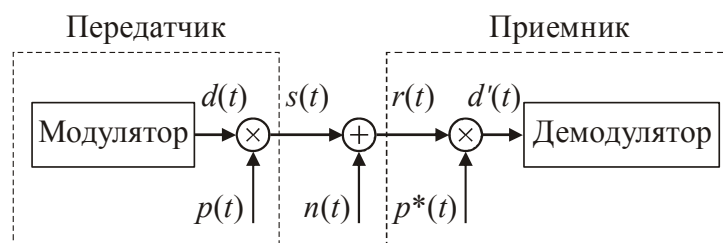


Рис. 5.3. Формирование и прием сигнала в системе с прямым расширением спектра

Расширяющая последовательность $p(t)$ представляет собой сигнал с фазовой модуляцией по псевдослучайному закону. Символы этого расширяющего сигнала называются *чипами* (английский глагол *to chip* дословно переводится как «нарезать»; в данном случае символ оказывается «нарезанным» на чипы) и имеют длительность $T_c < T$. Обратная величина $R_c = 1/T_c$ называется *чиповой скоростью* (chip rate). Ширина спектра расширяющего сигнала пропорциональна чиповой скорости R_c .

При перемножении исходного модулированного сигнала и расширяющей последовательности согласно свойствам преобразования Фурье их спектры сворачиваются. Ширина получающегося спектра примерно равна сумме ширин спектров перемножаемых сигналов и при большом отношении T/T_c определяется шириной спектра расширяющего сигнала. Указанное отношение называется *коэффициентом расширения спектра*: $G = T/T_c = R_c/R_{\text{sym}}$.

Пример информационного сигнала $d(t)$, расширяющей последовательности $p(t)$ и сигнала с расширенным спектром $s(t)$ приведен на рис. 5.4 для модуляции ФМ-2 (она используется как для информационного сигнала, так и для расширяющей последовательности) и коэффициента расширения спектра $G = 8$.

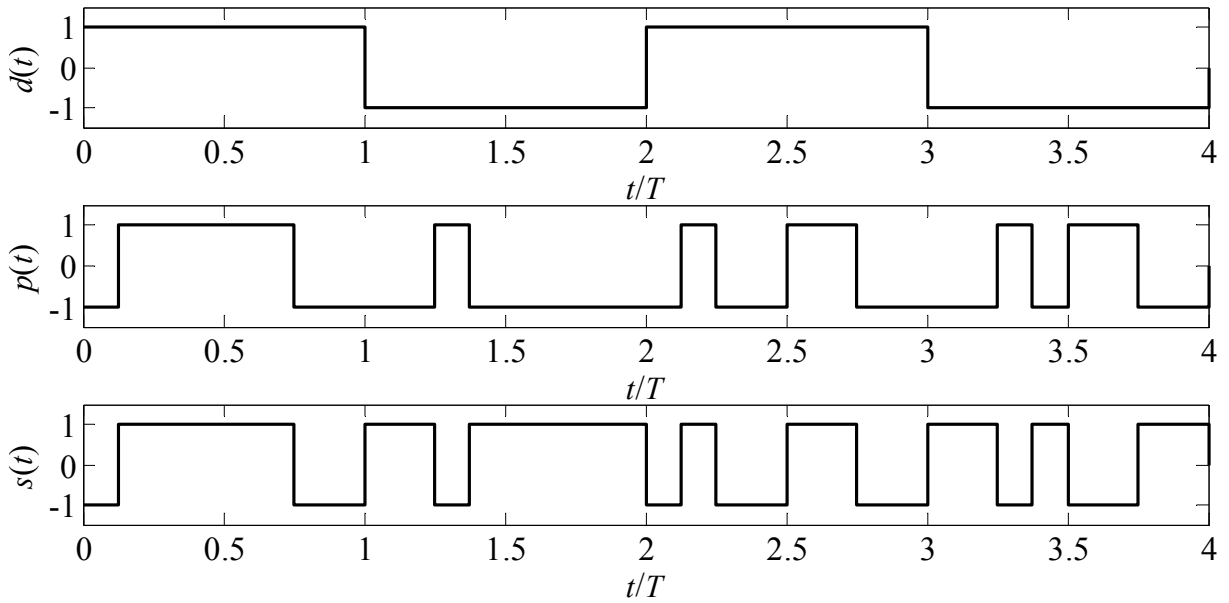


Рис. 5.4. Информационный сигнал (сверху), расширяющая последовательность (в центре) и сигнал с прямым расширением спектра (снизу)

В приемнике принятый сигнал $r(t)$ умножается на синхронизированную по времени и комплексно-сопряженную копию расширяющего сигнала: $p^*(t)$. В результате полезный сигнал восстанавливается (так как $p(t)p^*(t) = |p(t)|^2 = 1$), а шум оказывается умноженным на расширяющий сигнал:

$$\begin{aligned} d'(t) &= r(t)p^*(t) = (s(t) + n(t))p^*(t) = (d(t)p(t) + n(t))p^*(t) = \\ &= d(t)|p(t)|^2 + n(t)p^*(t) = d(t) + n(t)p^*(t). \end{aligned}$$

Однако умножение шума на расширяющую последовательность не меняет его свойства: так как не происходит нелинейного преобразования, он остается гауссовым, а поскольку преобразование является безынерционным, сохраняется дельта-коррелированность белого шума.

Таким образом, после умножения на комплексно-сопряженную расширяющую последовательность в приемнике полезный сигнал восстанавливается, а шум не меняет своих свойств. Поэтому процессы расширения и обратного сужения спектра никак не влияют на помехоустойчивость системы — она определяется только используемым видом модуляции. Здесь, однако, следует сделать одну оговорку — помехоустойчивость не меняется, если в качестве меры отношения сигнал/шум используется отношение E_b/N_0 . Что же касается отношения сигнал/шум по мощности, оно в результате расширения спектра уменьшается в G раз, так как мощность сигнала не меняется, а мощность шума в полосе сигнала увеличивается в G раз вследствие расширения этой полосы.

В зависимости от соотношения периода повторения расширяющей последовательности T_p и длительности символа различают следующие случаи:

- *длинные* расширяющие последовательности, $T_p > T$. Этот вариант обеспечивает более равномерный спектр и лучшую структурную скрытность, но он сложен в реализации, так как в приемнике требуется отдельная система для временной синхронизации расширяющей последовательности. Длинные расширяющие последовательности используются в сотовых системах связи с кодовым разделением каналов (CDMA). Например, в системе cdma2000 период расширяющей последовательности равен $2^{42} - 1$ чипов, что при чиповой скорости 1,2288 Мчип/с составляет более 40 сут;
- *короткие* расширяющие последовательности, $T_p \leq T$. В спектре сигнала в данном случае начинают проявляться регулярности, имеющиеся в коротком периоде повторения расширяющего сигнала, а структурная скрытность практически исчезает. Достоинством этого варианта является простота реализации. Наибольшее практическое распространение получил вариант, когда $T_p = T$ (т. е. $p(t - T) = p(t)$). При этом сигнал с линейной модуляцией и расширенным спектром можно представить следующим образом:

$$\begin{aligned} s(t) = d(t)p(t) &= \sum_{k=-\infty}^{\infty} a_k g(t - kT)p(t) = \sum_{k=-\infty}^{\infty} a_k g(t - kT)p(t - kT) = \\ &= \sum_{k=-\infty}^{\infty} a_k g'(t - kT), \end{aligned} \quad (5.3)$$

где $g'(t) = g(t)p(t)$, т. е. в данном случае расширение спектра свелось к модификации формы сигнальной посылки, в качестве которой теперь используется сигнал с большой базой. Вследствие этого отпадает необходимость в синхронизации генератора расширяющей последовательности в приемнике, нужна лишь обычная временная синхронизация демодулятора. Кроме того, блок сужения спектра можно объединить с демодулятором и реализовать на основе СФ.

Примером систем с прямым расширением спектра и короткой расширяющей последовательностью являются беспроводные сети WiFi, работающие по стандарту IEEE 802.11b. Сигнальная посылка в этой системе представляет собой 11-элементный код Баркера (рис. 5.5, а), так что коэффициент расширения спектра $G = 11$.

Коды Баркера, как известно, имеют идеальную (обеспечивающую минимально возможный уровень боковых лепестков) корреляционную функцию. КФ 11-элементного кода Баркера показана на рис. 5.5, б. Боковые лепестки в данном случае имеют отрицательную полярность и уровень 1/11 от максимума.

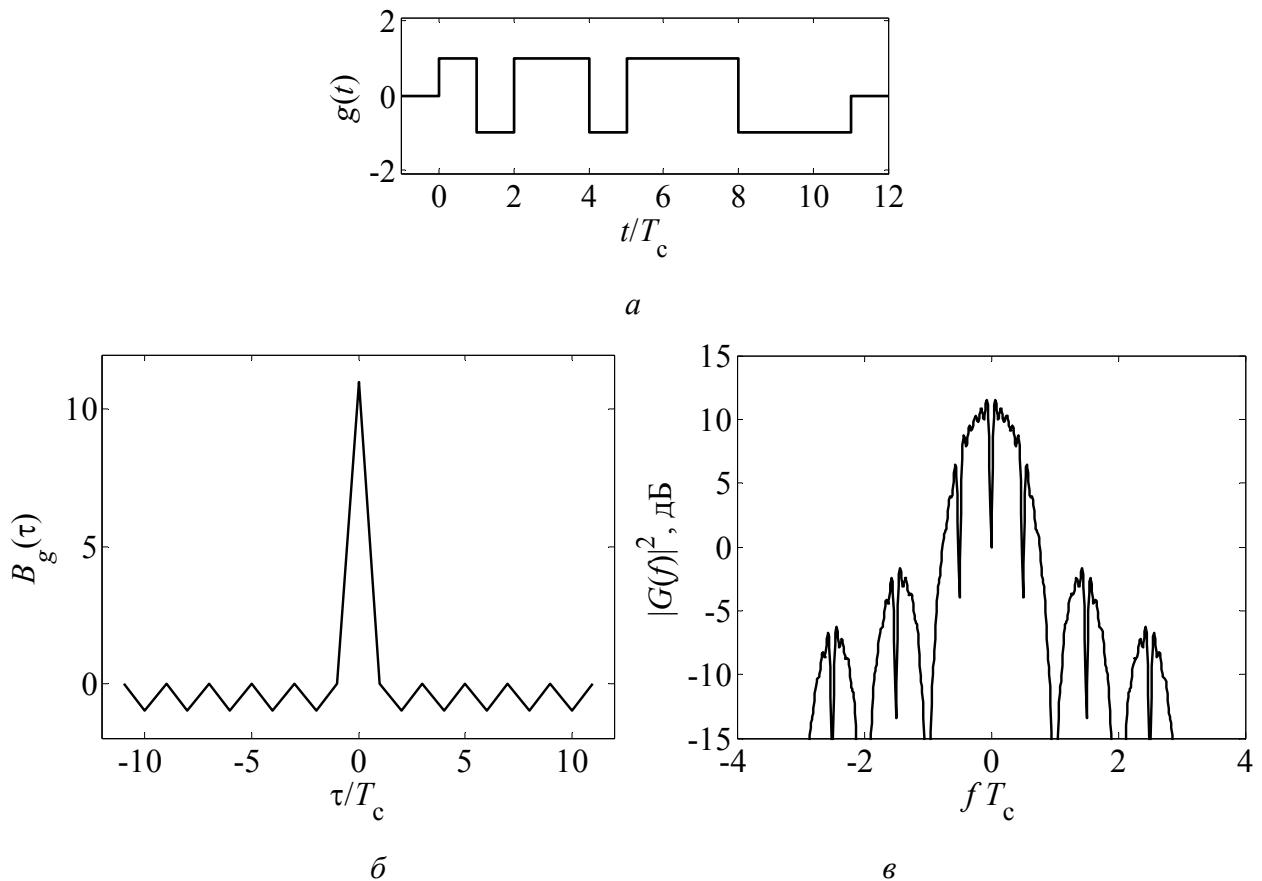


Рис. 5.5. 11-элементный код Баркера (а), его корреляционная функция (б) и энергетический спектр (в)

Рис. 5.5, в, где показан энергетический спектр данной сигнальной посылки, демонстрирует неравномерность спектра, возникающую при использовании коротких расширяющих последовательностей. В данном случае наблюдаются характерные узкие провалы спектра на нулевой частоте и ряде других частот.

5.2.2. Подавление узкополосной помехи

Механизм подавления узкополосной помехи в системах с прямым расширением спектра заключается в следующем. При умножении принятого сигнала на расширяющую последовательность спектр полезного сигнала сужается, так как расширяющие последовательности в приемнике и передатчике синхронизированы друг с другом. Спектр помехи, напротив, расширяется, так как она не коррелирована с расширяющей последовательностью (рис. 5.6).

Далее следует демодулятор, который содержит, помимо прочего, фильтр нижних частот (реализованный в виде СФ либо коррелятора) с полосой

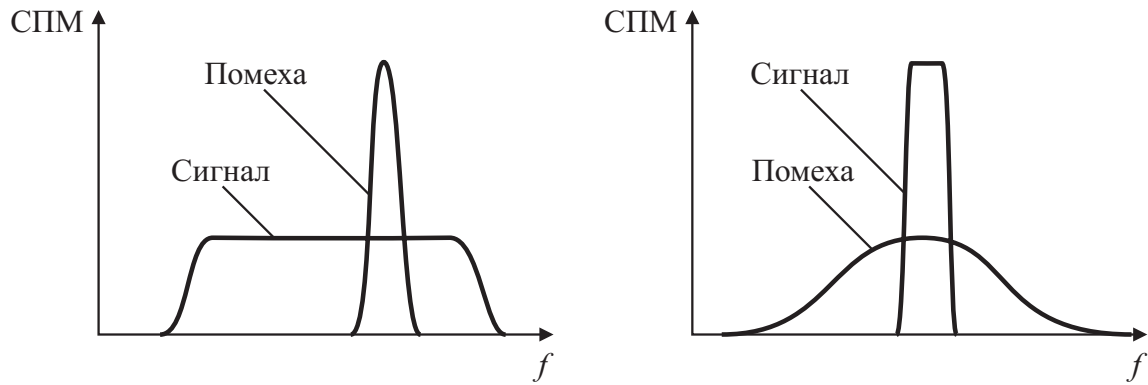


Рис. 5.6. Спектры сигнала с прямым расширением спектра и узкополосной помехи на входе приемника (слева) и после умножения на расширяющую последовательность (справа)

пропускания, равной символьной скорости $R_{\text{sym}} = 1/T$. Большая часть «размазанного» спектра помехи оказывается вне полосы пропускания этого фильтра, в результате мощность помехи ослабляется примерно в G раз.

5.2.3. Подавление широкополосной помехи

В качестве примера воздействия широкополосной помехи рассмотрим помеху импульсную, моделируя ее как дополнительный белый гауссов шум, действующий не во все моменты времени. Введем следующие обозначения:

- μ — доля времени, пораженного импульсным шумом;
- I_0 — усредненная по времени СПМ импульсного шума.

Тогда СПМ импульсного шума во время действия импульса будет равна I_0/μ .

Будем также считать, что символы сигнала поражаются или не поражаются импульсной помехой только целиком. Тогда с вероятностью $(1 - \mu)$ на систему действует только фоновый АБГШ с СПМ, равной N_0 , а с вероятностью μ к нему добавляется дополнительный белый шум с СПМ, равной I_0/μ . Если в системе используется модуляция ФМ-2, для средней вероятности битовой ошибки получается следующая формула:

$$P_b = (1 - \mu)Q\left(\sqrt{\frac{2E_b}{N_0}}\right) + \mu Q\left(\sqrt{\frac{2E_b}{N_0 + I_0/\mu}}\right). \quad (5.4)$$

Если доминирующим является импульсный шум, величиной N_0 в (5.4) можно пренебречь:

$$P_b \approx \mu Q\left(\sqrt{\frac{2\mu E_b}{I_0}}\right). \quad (5.5)$$

Семейство кривых, полученных по формуле (5.5) при разных μ , показано на рис. 5.7. По горизонтальной оси отложено отношение сигнал/помеха — E_b/I_0 .

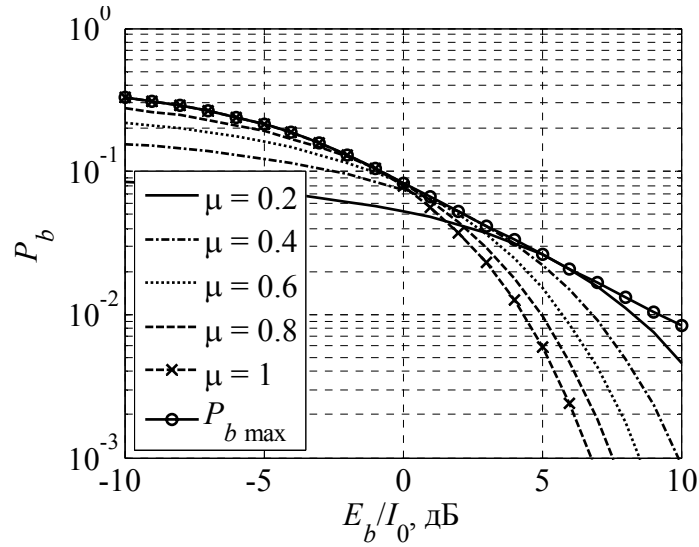


Рис. 5.7. Помехоустойчивость системы с ФМ-2 при воздействии импульсной помехи

Штриховая кривая с крестиками соответствует случаю $\mu = 1$, т. е. помехе, действующей непрерывно. Из графиков видно, что при снижении значения μ вероятность ошибки возрастает при большом отношении E_b/I_0 , но становится меньше при малом отношении E_b/I_0 . Поэтому для каждого значения E_b/I_0 существует наихудшее значение μ_0 доли пораженного помехой времени, при котором вероятность ошибки оказывается наибольшей. Расчеты показывают, что

$$\mu_0 \approx \begin{cases} \frac{0,7}{E_b/I_0}, & E_b/I_0 > 0,7, \\ 1, & E_b/I_0 \leq 0,7. \end{cases} \quad (5.6)$$

Само же наибольшее значение вероятности ошибки, получаемое при наихудшем $\mu = \mu_0$, рассчитывается подстановкой (5.6) в (5.5):

$$P_{b \max} \approx \begin{cases} \frac{0,083}{(E_b/I_0)}, & E_b/I_0 > 0,7, \\ Q\left(\sqrt{\frac{2E_b}{I_0}}\right), & E_b/I_0 \leq 0,7. \end{cases}$$

Вероятность $P_{b \max}$ дает верхнюю огибающую семейства графиков на рис. 5.7 (сплошная линия с кружочками).

5.2.4. Устойчивость к многолучевому распространению радиоволн

Если задержка между лучами принимаемого сигнала превышает длительность чипа T_c , то правильная временная синхронизация расширяющей последовательности в приемнике может наблюдаться только для одного из

поступающих на вход лучей. Остальные лучи будут умножаться на несинхронизированные расширяющие последовательности и в результате будут восприниматься приемником просто как дополнительный шум. Таким образом, благодаря расширению спектра приемник без каких-либо специальных мер может осуществлять прием сигнала, используя один из имеющихся лучей.

Однако в данной ситуации, если разброс времени прихода лучей не превышает длительности символа, можно собрать и использовать энергию всех имеющихся лучей, трактуя их как ветви разнесенного приема и организовав оптимальное сложение этих ветвей. Такой подход называется RAKE-приемником (английское слово *rake* дословно переводится как «грабли»; имеется в виду, что вид ИХ многолучевого канала связи, а также вид сигнала после СФ напоминают этот инструмент).

Объяснить принцип работы RAKE-приемника проще для случая короткой расширяющей последовательности, когда сигнал описывается формулой (5.3). Пусть многолучевой канал связи имеет ИХ вида

$$h(t) = \sum_k \dot{r}_k \delta(t - \tau_k),$$

где \dot{r}_k — коэффициент передачи k -го луча; τ_k — его задержка. После прохождения через такой канал связи сигнальная посылка принимает форму

$$g_h(t) = \sum_k \dot{r}_k g'(t - \tau_k),$$

где $g'(t)$ — сигнальная посылка в виде широкополосного сигнала (см. (5.3)).

Импульсную характеристику фильтра, согласованного с этой искаженной каналом сигнальной посылкой, можно представить следующим образом (необходимость введения задержки для получения физически реализуемой системы пока что проигнорируем):

$$g_h^*(-t) = \sum_k \dot{r}_k^* g'^*(-t - \tau_k) = \sum_k \dot{r}_k^* g'^*(-(t + \tau_k)).$$

Так как $g'^*(-t)$ — импульсная характеристика фильтра, согласованного с сигнальной посылкой, ИХ искомого фильтра представляет собой линейную комбинацию сдвинутых по времени ИХ фильтра, согласованного с исходной (не искаженной каналом) сигнальной посылкой $g'(t)$. Отсюда следует, что требуемый фильтр можно представить в виде каскадного соединения фильтра, согласованного с сигнальной посылкой, и RAKE-структуры, реализующей линейную комбинацию задержанных сигналов с необходимыми весовыми коэффициентами. Импульсная характеристика этой структуры будет иметь вид

$$h_{\text{RAKE}}(t) = \sum_k \dot{r}_k^* \delta(t + \tau_k).$$

В полученной формуле дельта-функции сдвинуты *влево* относительно точки начала отсчета времени, поэтому такая структура физически нереализуема. Чтобы решить эту проблему, задержим ИХ на время τ_{\max} , равное максимальной задержке приходящих на вход лучей: $\tau_{\max} = \max\{\tau_k\}$:

$$h_{\text{RAKE}}(t - \tau_{\max}) = \sum_k \dot{r}_k^* \delta(t - (\tau_{\max} - \tau_k)).$$

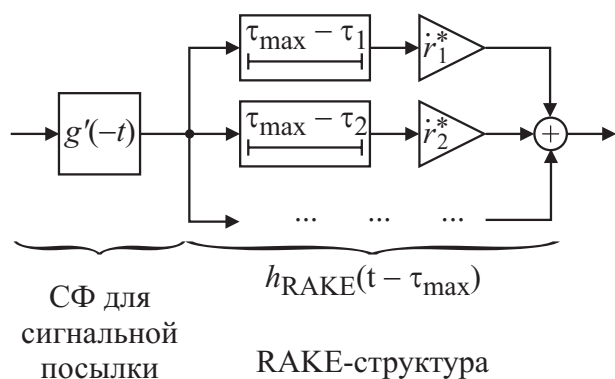


Рис. 5.8. RAKE-приемник

Схема полученной структуры, представляющей собой каскадное соединение фильтра, согласованного с сигнальной посылкой, и собственно RAKE-структуры, осуществляющей оптимальное сложение задержанных лучей, показана на рис. 5.8.

Разумеется, возможна реализация аналогичной структуры и при использовании длинных расши-

ряющих последовательностей. В этом случае каждому имеющемуся лучу соответствует отдельный канал RAKE-приемника, эти каналы называются «пальцами» (*fingers*). В каждом канале входной сигнал умножается на расширяющую последовательность с соответствующим сдвигом, после чего обрабатывается согласованным с (узкополосной) сигнальной посылкой фильтром или коррелятором. Выходные сигналы СФ или корреляторов задерживаются и суммируются так же, как показано на рис. 5.8.

Следует отметить, что такой способ приема возможен только для сигнала с расширенным спектром. Если попытаться использовать данный подход для узкополосного сигнала, после реализации СФ, согласованного с сигнальной посылкой, искаженной каналом связи, получится очень сильная межсимвольная интерференция. При использовании же широкополосного сигнала КФ сигнальной посылки представляет собой узкий пик, а КФ посылки, искаженной многолучевым каналом — набор узких пиков. Если эти пики не попадают на задержки, кратные длительности символа, заметной МСИ не возникнет.

Чтобы пояснить изложенное, рассмотрим в качестве примера систему, использующую сигнальную посылку в виде 11-элементного кода Баркера (см. рис. 5.5), и канал связи, содержащий 3 луча со следующими задержками и коэффициентами передачи:

- $\tau_1 = 0, r_1 = 1$;
- $\tau_2 = 5$ чипов, $r_2 = 0,6$;
- $\tau_3 = 7$ чипов, $r_3 = -0,4$.

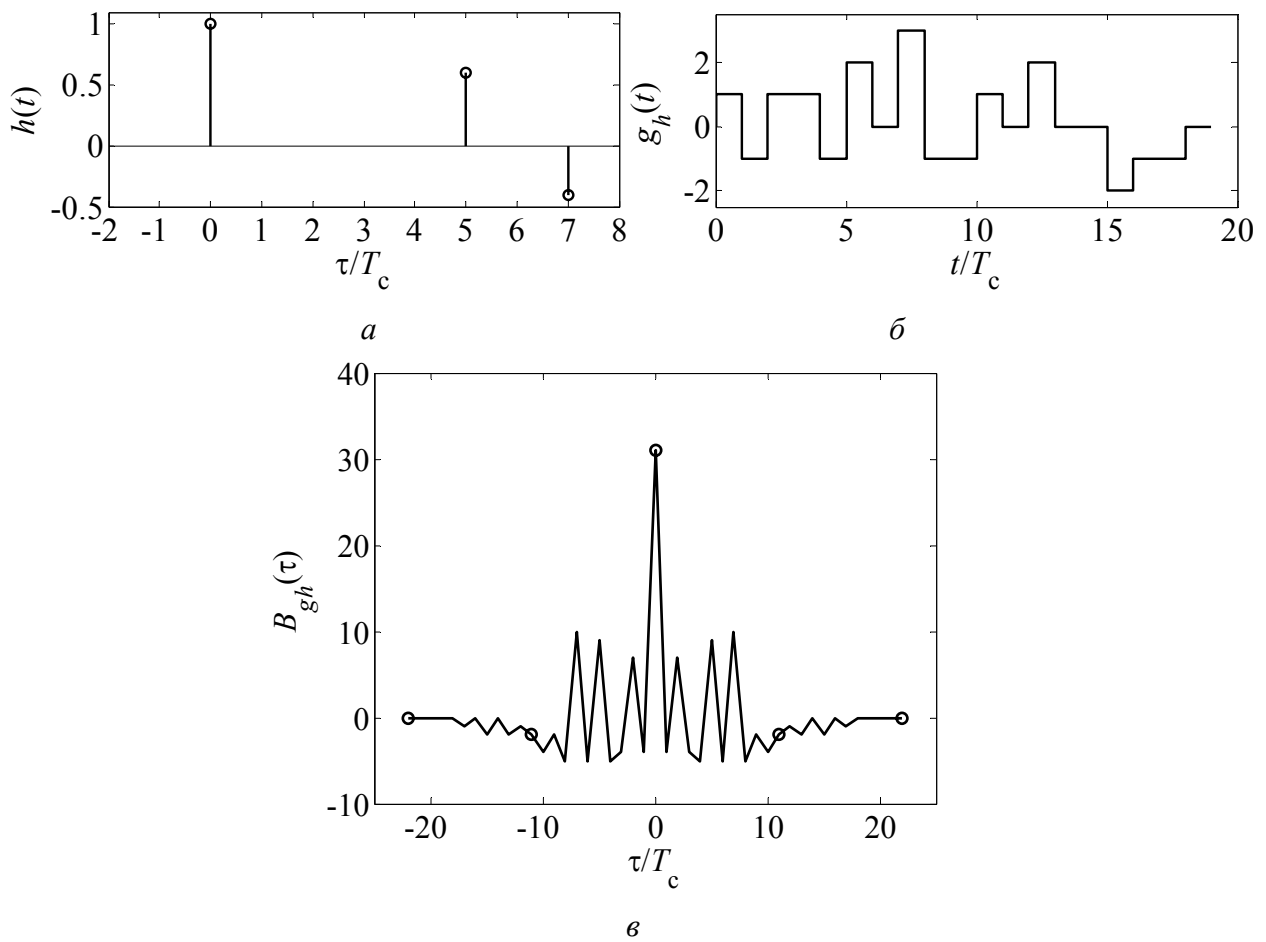


Рис. 5.9. Прием 11-элементного кода Баркера в трехлучевом канале: *а* — ИХ канала связи; *б* — сигнальная посылка, искаженная каналом; *в* — КФ искаженной сигнальной посылки

На рис. 5.9 показаны ИХ канала связи, сигнальная посылка, искаженная данным каналом связи, и ее КФ. На графике КФ (рис. 5.9, в) кружками показаны моменты взятия отсчетов сигнала после РАКЕ-приемника. Видно, что пики КФ расположены в пределах длительности символа, а МСИ, вызванная боковыми лепестками КФ (ее уровень определяется значениями КФ для задержки, равной длительности символа, т. е. 11 чипам), незначительна.

5.2.5. Реализация множественного доступа

Совместное использование частотно-временного ресурса в случае прямого расширения спектра реализуется с помощью разных расширяющих последовательностей. Это называется множественным доступом с кодовым разделением каналов. Подробнее данный вопрос рассматривается в разд. 6, здесь отметим только, что в зависимости от наличия или отсутствия взаимной временной синхронизации может наблюдаться либо полное отсутствие взаимных помех, либо плавное нарастание их уровня пропорционально числу одновременно работающих систем.

5.3. Псевдослучайная перестройка рабочей частоты

5.3.1. Принципы реализации ППРЧ

Смысл данного метода расширения спектра очевиден из его названия: несущая частота сигнала время от времени меняется. Интервал между *скачками* (англ. термин — *hop*) частоты будем обозначать как T_h , обратная величина называется *частотой скачков*: $F_h = 1/T_h$.

Как правило, ППРЧ реализуется в блоке преобразования частоты, который осуществляет перенос комплексной огибающей сигнала (или сигнала, сформированного на фиксированной промежуточной частоте) на несущую частоту. Для этого используется перестраиваемый генератор (как правило, синтезатор частоты), перестройка которого осуществляется с помощью псевдослучайного расширяющего кода, задающего в данном случае последовательность используемых частот. Общая структура передатчика и приемника системы с ППРЧ показана на рис. 5.10.

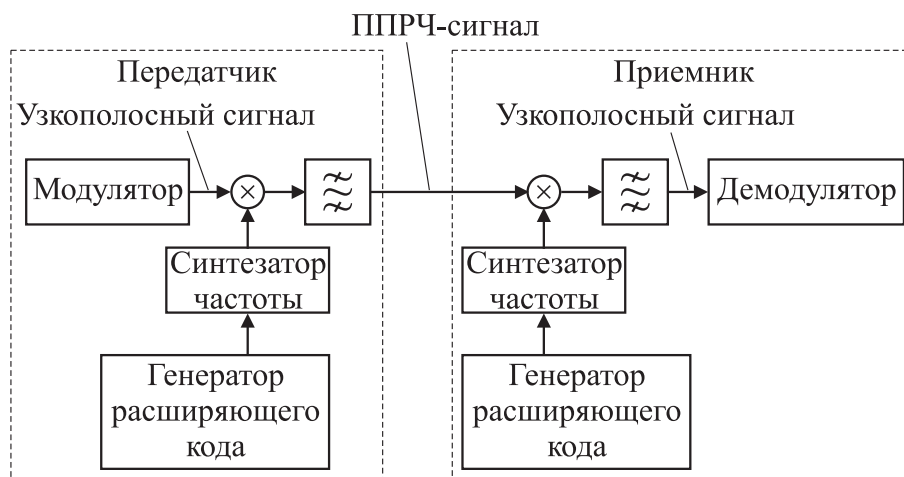


Рис. 5.10. Формирование и прием сигнала с ППРЧ

В зависимости от соотношения между периодом скачков T_h и длительностью символа T различают *быструю* и *медленную* ППРЧ:

- Медленная ППРЧ (Slow FHSS): $T_h \geq T$, т. е. на одной несущей частоте передается несколько символов подряд.
- Быстрая ППРЧ (Fast FHSS): $T_h < T$, т. е. несущая частота меняется в процессе передачи символа.

5.3.2. Медленная ППРЧ

Медленная ППРЧ проще в реализации, так как, фактически, здесь используется обычный приемопередатчик, время от времени перестраиваемый по частоте. В принципе, вид модуляции в данном случае может быть любым, так как при редких скачках частоты есть возможность после каждого скачка

вставлять в сигнал служебные фрагменты для фазовой синхронизации приемника, а значит, можно реализовать когерентный прием. Однако на практике чаще всего в ППРЧ-системах используется частотная модуляция.

На рис. 5.11 показано использование частотно-временной плоскости при медленной ППРЧ ($T_h = 4T$) с применением четырех несущих частот и модуляции ЧМ-2. Таким образом, всего для передачи сигнала используется 8 частот, которые на рис. 5.11 обозначены как f_{nk} , где n — номер несущей частоты ($n = 1, \dots, 4$), а k — значение передаваемого бита ($k = 0, 1$). Штриховые прямоугольники показывают зоны на частотно-временной плоскости, в пределах которых должен располагаться сигнал при данном значении несущей частоты и различных передаваемых данных.

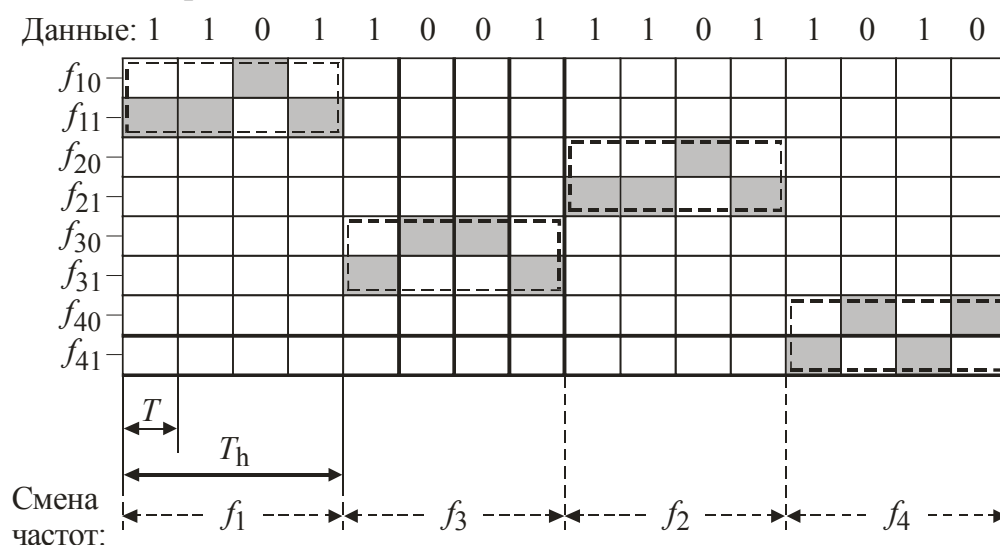


Рис. 5.11. Медленная ППРЧ

Примером системы с медленной ППРЧ является Bluetooth (стандарт IEEE 802.15.1). В этой системе используется 79 несущих частот, образующих равномерную сетку с шагом 1 МГц. Модуляция — ЧМ-2 с непрерывной фазой и гауссовым сглаживанием, длительность символа составляет $T = 1$ мкс (скорость передачи данных 1 Мбит/с). Интервал между скачками частоты $T_h = 625$ мкс (1600 скачков в секунду).

5.3.3. Быстрая ППРЧ

Быстрая ППРЧ обладает рядом преимуществ перед медленной, но реализуется существенно сложнее, так как при приеме необходимо собрать вместе кусочки символа, разбросанные по разным частотам. Для когерентного приема необходимо, чтобы при смене несущей частоты жестко выдерживались заданные фазовые соотношения между фрагментами сигнала. На практике реализовать это можно только при цифровом формировании ППРЧ-сигнала, а такой вариант практически эквивалентен прямому расширению

спектра. При формировании частотных скачков за счет управления синтезатором частоты нет возможности контролировать фазовые соотношения между фрагментами символа, поэтому практически единственным видом модуляции, который может использоваться в данном случае, является ЧМ с некогерентным приемом и с некогерентным сложением фрагментов символа.

На рис. 5.12 показано использование частотно-временной плоскости при быстрой ППРЧ ($T_h = T/4$) с применением четырех несущих частот и модуляции ЧМ-2. Обозначения частот построены так же, как и на рис. 5.11, однако эти частоты на рисунке расположены в другой последовательности. Штриховые прямоугольники показывают зоны на частотно-временной плоскости, в пределах которых должен располагаться сигнал при данном значении передаваемого бита и различных несущих частотах.

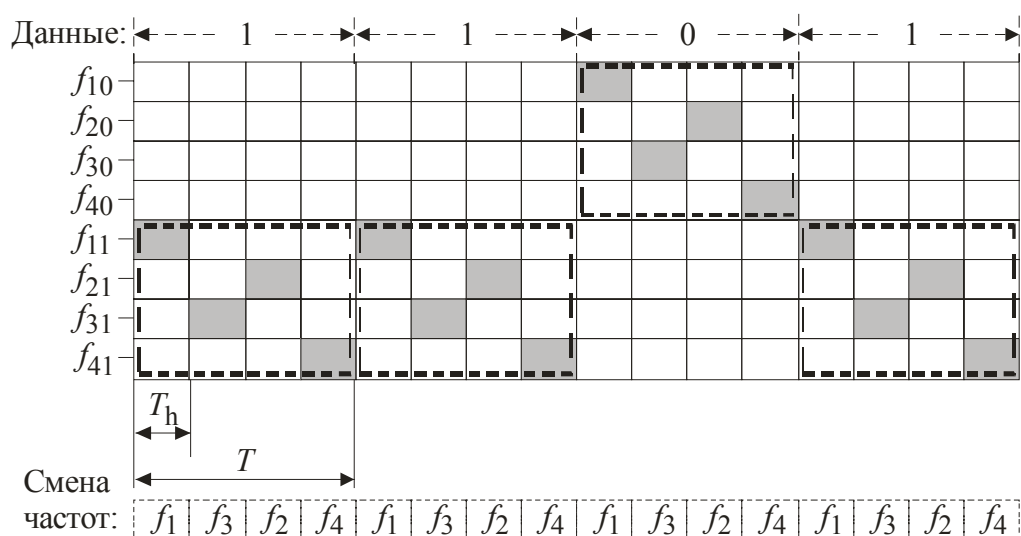


Рис. 5.12. Быстрая ППРЧ

Из-за сложности реализации быстрая ППРЧ используется только в системах связи военного и специального назначения.

5.3.4. Подавление узкополосной помехи

Механизмы реализации преимуществ от расширенного спектра в системах с быстрой и медленной ППРЧ различны.

За счет перестройки частоты узкополосная помеха, которая затрагивает только часть используемых системой несущих частот, будет влиять на прием сигнала только на тех интервалах времени, когда система использует соответствующие частоты. Таким образом, узкополосная помеха за счет псевдослучайной перестройки частоты преобразуется в помеху импульсную. В системах с медленной ППРЧ оказываются поражены помехой блоки (пакеты) символов, передаваемые на частотах, пораженных помехой. Это вызывает группирование ошибок, и для борьбы с ними необходимо применять помехоустойчивое кодирование в сочетании с перемежением (о перемежении см.

разд. 4). При использовании быстрой ППРЧ символ «размазан» по разным частотам, поэтому поражается помехой лишь его часть. Таким образом, быстрая ППРЧ обладает устойчивостью к узкополосной помехе даже без использования помехоустойчивого кодирования.

Получим формулы для средней вероятности ошибки в системе с модуляцией ЧМ-2 и некогерентным приемом при воздействии помехи частичной полосы, занимающей долю μ общей полосы частот системы. Будем считать, что символы поражаются или не поражаются узкополосной помехой только целиком, что на практике наблюдается при медленной ППРЧ. Использование подхода, аналогичного примененному ранее при анализе системы с прямым расширением спектра при воздействии импульсной помехи, дает следующий результат:

$$P_b = \frac{1-\mu}{2} \exp\left(-\frac{E_b}{2N_0}\right) + \frac{\mu}{2} \exp\left(-\frac{E_b}{2(N_0 + I_0/\mu)}\right), \quad (5.7)$$

где I_0 — СПМ помехи, усредненная в пределах полной полосы частот системы (СПМ в полосе воздействия помехи при этом будет равна I_0/μ).

Если доминирующей является помеха частичной полосы, величиной N_0 в (5.7) можно пренебречь:

$$P_b \approx \frac{\mu}{2} \exp\left(-\frac{\mu E_b}{2I_0}\right). \quad (5.8)$$

Семейство кривых, полученных по формуле (5.8) при разных μ , показано на рис. 5.13. По горизонтальной оси отложено отношение «сигнал/помеха» — E_b/I_0 .

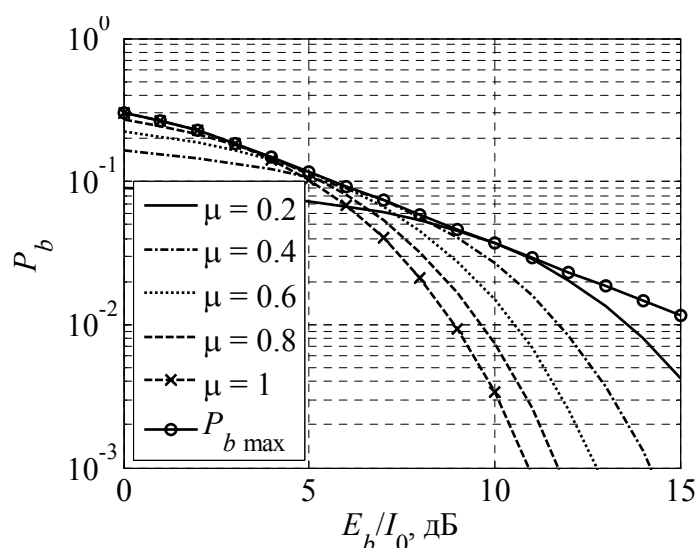


Рис. 5.13. Помехоустойчивость ППРЧ-системы с некогерентной ЧМ-2 при воздействии помехи частичной полосы

Штриховая кривая с крестиками соответствует случаю $\mu = 1$, т. е. помехе, перекрывающей всю полосу частот системы. Из графиков видно, что при снижении значения μ вероятность ошибки возрастает при большом отношении сигнал/помеха и снижается при малом отношении сигнал/помеха. Поэтому для каждого значения E_b/I_0 существует наихудшее значение μ_0 доли пораженной помехой полосы, при котором вероятность ошибки оказывается наибольшей. Расчеты показывают, что

$$\mu_0 = \begin{cases} \frac{2}{(E_b/I_0)}, & E_b/I_0 > 2, \\ 1, & E_b/I_0 \leq 2. \end{cases} \quad (5.9)$$

Само же наибольшее значение вероятности ошибки, получаемое при наихудшем $\mu = \mu_0$, рассчитывается подстановкой (5.9) в (5.8):

$$P_{b \max} = \begin{cases} \frac{e^{-1}}{(E_b/I_0)}, & E_b/I_0 > 2, \\ \frac{1}{2} \exp\left(-\frac{E_b}{2I_0}\right), & E_b/I_0 \leq 2. \end{cases}$$

Вероятность $P_{b \max}$ дает верхнюю огибающую семейства графиков на рис. 5.13 (сплошная линия с кружочками).

5.3.5. Устойчивость к замираниям радиосигнала

Ввиду расширения спектра сигнала замираниями поражаются в каждый момент лишь некоторые из несущих частот. При медленной ППРЧ это приведет к тому, что некоторые пакеты сигнала (попавшие на частоты, где в данный момент произошло замирание) будут приняты с повышенной вероятностью ошибки. Для борьбы с этими группирующимися ошибками необходимо использование помехоустойчивого кодирования в сочетании с перемежением. При быстрой ППРЧ замиранию окажется подвержена лишь часть символа, что «усреднит» влияние замираний даже без использования помехоустойчивого кодирования. Можно считать, что механизм устойчивости к замираниям оказывается аналогичен механизму устойчивости к узкополосной помехе.

5.3.6. Устойчивость к многолучевому распространению радиоволн

Если задержка τ между принимаемыми лучами превышает T_h , приемник к моменту поступления на вход задержанного сигнала успевает перестроиться на новую частоту, так что интерференция между лучами не возникает. В случае быстрой ППРЧ вероятность выполнения данного условия ($\tau \geq T_h$), очевидно, намного больше, чем при медленной ППРЧ. Если же $\tau < T_h$, ре-

зультат зависит от типа используемой ППРЧ. При медленной ППРЧ все происходит так же, как при отсутствии перестройки частоты: сигнал испытывает селективные или неселективные замирания в зависимости от того, как соотносятся величина $1/\tau$ и ширина спектра излучаемого (на одной несущей частоте) сигнала (см. разд. 4). При быстрой ППРЧ данное условие означает, что все лучи приходят в течение передачи «кусочка» символа на одной частоте. В результате замирание для этого «кусочка» окажется неселективным, но результирующий коэффициент передачи для разных «кусочков» символа может оказаться различным. При сложении фрагментов символа коэффициент передачи усредняется по частоте, что фактически означает использование частотного разнесения. Таким образом, система с быстрой ППРЧ обладает большей устойчивостью к многолучевому распространению и замираниям, чем система с медленной ППРЧ.

5.3.7. Реализация множественного доступа

При использовании ППРЧ одновременная работа нескольких систем в общем частотном диапазоне реализуется с помощью разных последовательностей перестройки частоты. Если число используемых несущих частот существенно превышает число одновременно работающих радиолиний, вероятность их одновременной работы на одной частоте (это называется *коллизией*, т. е. «столкновением») оказывается невелика. При медленной ППРЧ отдельные пакеты данных окажутся потерянными в результате коллизий. Эта проблема решается за счет механизмов подтверждения доставки и повторной передачи пакетов. При быстрой ППРЧ взаимные помехи затрагивают лишь фрагменты символов, так что помехоустойчивость снижается незначительно.

Указанным образом реализовано совместное использование частотно-временного ресурса в упоминавшейся ранее системе Bluetooth. В одной и той же области пространства может существовать несколько так называемых *пикосетей* Bluetooth. В пределах пикосети используется временное разделение каналов (о разделении каналов см. разд. 6), а разные пикосети используют разные последовательности перестройки частоты. Одно из устройств пикосети является *ведущим*, и псевдослучайный код, задающий последовательность использования частот, генерируется из уникального сетевого адреса этого устройства.

6. МНОЖЕСТВЕННЫЙ ДОСТУП

6.1. Структуры многопользовательских систем

Можно выделить две основные архитектуры многопользовательских систем связи:

- системы с базовыми станциями;
- системы без базовых станций (ad hoc networks).

В системах с базовыми станциями пользователи обмениваются сигналами только с базовыми станциями, а не непосредственно между собой. Базовые станции не только играют роль ретрансляторов, но и координируют работу всей сети. Примерами таких систем являются сети сотовой связи, транкинговые системы и т. п.

В системах без базовой станции глобальная координация работы сети отсутствует, пользователи обмениваются информацией непосредственно друг с другом, возможно, с участием других пользовательских терминалов в качестве ретрансляторов.

В сетях с базовыми станциями существенно различаются по своим свойствам направления передачи «вниз» (от базовой станции к пользователям, downlink) и «вверх» (от пользователей к базовой станции, uplink) (рис. 6.1). Для разделения этих направлений (реализация одновременных (с точки зрения пользователя) передачи и приема называется *дуплексом*) используется частотный либо временной дуплекс (Frequency Division Duplex, FDD, или Time Division Duplex, TDD, соответственно).

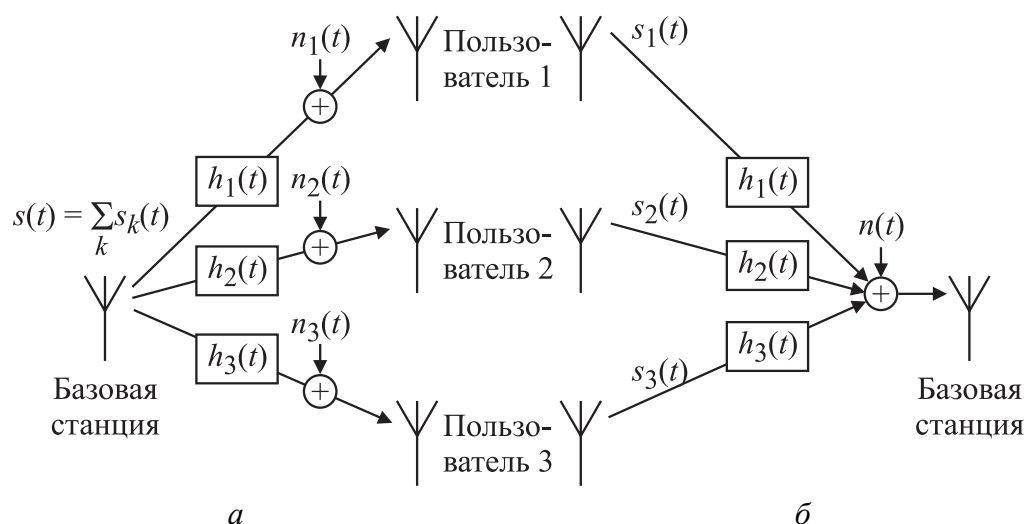


Рис. 6.1. Каналы «вниз» и «вверх» в многопользовательской системе связи с базовой станцией

Канал «вниз» (см. рис. 6.1, *а*) характеризуется следующими свойствами:

- так как групповой сигнал $s(t)$, представляющий собой сумму сигналов $s_k(t)$, адресованных всем пользователям, формируется базовой станцией, автоматически обеспечивается идеальная временная синхронизация;
- в точке приема «свой» сигнал и помехи от сигналов других пользователей искажаются каналом связи одинаково.

Канал «вверх» (см. рис. 6.1, *б*) обладает следующими свойствами:

- пользователи излучают свои сигналы $s_k(t)$ независимо, кроме того, они находятся на разных расстояниях от базовой станции, поэтому временная синхронизация в данном случае отсутствует. В принципе, за счет обратной связи (команд базовой станции, посылаемых пользовательским терминалам) можно в определенной степени «выровнять» моменты прихода разных сигналов, но это очень сильно усложняет систему;
- сигнал от каждого пользователя проходит через *свой* канал связи, поэтому все сигналы приобретают *разные* искажения.

В зависимости от способа организации доступа пользователей к общей среде передачи можно выделить 2 варианта:

- При непрерывной передаче информации с заданной скоростью в течение определенного промежутка времени используется технология *разделения каналов* (или *множественного доступа*, multiple access), когда каждому пользователю выделяется определенный фрагмент общего частотно-временного ресурса. К такому типу относятся, например, системы сотовой связи.
- Если информация для передачи возникает неравномерно, а требования к задержкам передачи не очень жесткие, применяется *случайный доступ* (random access) к среде передачи, когда пользователи используют среду передачи совместно, причем не координируя или почти не координируя свои действия. Доступ к среде передачи в этой ситуации не гарантирован, возможны конфликты доступа. Примером таких систем являются компьютерные сети.

Далее две указанные технологии будут рассмотрены более подробно.

6.2. Разделение каналов (множественный доступ)

При разделении каналов каждому пользователю выделяется определенная часть общего частотно-временного ресурса. Чтобы избежать взаимных помех, сигналы разных пользователей должны быть ортогональны друг другу. Таким образом, если рассматривается фрагмент частотно-временной плоскости длительностью Δt и с шириной полосы Δf , то максимальное количество пользователей определяется размерностью данного сигнального подпространства. Согласно теореме Котельникова, комплексная огибающая сигнала с полосой Δf может дискретизироваться с периодом $1/\Delta f$ без потери ин-

формации. За время Δt получается $\Delta f \Delta t$ отсчетов. Кроме того, отсчеты при дискретизации комплексной огибающей представляют собой комплексные числа, т. е. являются двумерными. Таким образом, размерность сигнального подпространства составляет в данном случае $2\Delta f \Delta t$. Это и есть максимальное число пользователей в системе. Однако достигнуть этого количества можно только в случае использования модуляции с вещественным созвездием (например, ФМ-2), при комплексных созвездиях предельное число пользователей составляет $\Delta f \Delta t$.

В соответствии с принципом выделения фрагментов частотно-временного ресурса различают частотное, временное и кодовое разделение каналов. Рассмотрим основные черты этих методов подробнее.

6.2.1. Частотное разделение каналов

При частотном разделении каналов (ЧРК; англ. термин — Frequency Division Multiple Access (FDMA)) каждому пользователю выделяется определенная *полоса частот*. Основные черты данного метода:

- необходимы частотные промежутки между каналами для защиты от факторов, вызывающих взаимные помехи (нестабильность генераторов, неидеальность фильтрации помех по соседним каналам, доплеровский сдвиг частоты и т. п.). Это уменьшает достижимое число пользователей;
- приемник должен быть перестраиваемым по частоте, что снижает технологичность оборудования;
- сложно реализовать гибкое распределение ресурса между пользователями (для выделения одному пользователю нескольких частотных полос приемник должен быть многоканальным, что сильно усложняет его);
- сигнал каждого пользователя является относительно узкополосным, что уменьшает вероятность возникновения частотно-селективных искажений;
- мощность излучаемого сигнала не меняется во времени;
- передача является непрерывной, поэтому системы синхронизации и эквалайзеры приемника должны только отслеживать относительно медленные изменения параметров канала — это проще, чем устанавливать синхронизацию каждый раз «с нуля».

В цифровой связи данный способ разделения каналов, как правило, не применяется в чистом виде (исключением являются системы цифрового радио- и телевизионного вещания), а используется в сочетании с другими методами. Например, в системе сотовой связи GSM используется комбинация частотного и временного разделения каналов.

В последнее время частотное разделение каналов начинает использоваться в сочетании с технологией OFDM, при этом разным пользователям выделяются разные группы поднесущих OFDM-сигнала. Данный метод получил аббревиатуру OFDMA (Orthogonal Frequency Division Multiple Access).

6.2.2. Временное разделение каналов

При временном разделении каналов (ВРК; англ. термин — Time Division Multiple Access (TDMA)) каждому пользователю выделяются определенные *временные интервалы*. Основные черты данного метода:

- необходимы временные промежутки между интервалами, выделенными разным пользователям, для защиты от факторов, вызывающих взаимные помехи (разброс расстояний от пользователей до базовой станции и вследствие этого разброс задержек прихода сигналов, неидеальная временная синхронизация и т. п.). Это уменьшает достижимое число пользователей;
- не требуется перестройка приемника по частоте, что упрощает аппаратуру;
- выделение одному пользователю нескольких временных интервалов реализуется без изменения аппаратной части радиочастотного тракта приемника;
- сигнал каждого пользователя занимает всю отведенную системе полосу частот, что увеличивает вероятность возникновения частотно-селективных искажений;
- передача является прерывистой, т. е. мощность излучаемого сигнала меняется во времени. Так как помехоустойчивость определяется в конечном счете средней мощностью, приходится увеличивать мощность во время передачи пакета данных (по сравнению со случаем, когда передача ведется непрерывно);
- из-за прерывистого характера передачи необходима буферизация поступающих на вход передатчика данных;
- необходима временная синхронизация между пользователями, поэтому работа системы должна координироваться базовой станцией. Это приводит к необходимости передавать соответствующую служебную информацию и, вследствие этого, к снижению общей пропускной способности системы;
- реализовать временную синхронизацию в канале «вверх» сложно, так как сигналы от разных пользователей проходят через разные каналы и приобретают разные задержки;
- из-за прерывистой передачи характеристики канала на приемной стороне каждый раз нужно оценивать *заново*, это сложнее, чем отслеживать их постепенные медленные изменения.

Как уже отмечалось, данный метод используется, в частности, в сетях сотовой связи GSM (в комбинации с частотным разделением каналов) — каждому пользователю выделяется полоса частот и набор временных интервалов.

6.2.3. Кодовое разделение каналов

При кодовом разделении каналов (КРК; англ. термин — Code Division Multiple Access (CDMA)) каждый пользователь использует частотно-временной ресурс целиком, следовательно, в данном случае должны применяться технологии расширения спектра (см. разд. 5). Разным пользователям назначаются различные расширяющие последовательности, за счет чего обеспечивается ортогональность их сигналов и возможность разделения каналов на приемной стороне.

Являясь системой с расширением спектра, система с CDMA может использовать все преимущества такой технологии (см. разд. 5).

Возможности использования CDMA существенно различаются для направлений «вверх» и «вниз».

В направлении «вниз» используется *синхронный* вариант CDMA (S-CDMA). Его легко реализовать, так как групповой сигнал формируется базовой станцией и, следовательно, автоматически обеспечивается жесткая временная синхронизация между пользователями. В этом случае в качестве расширяющих последовательностей используются наборы ортогональных сигналов. Для бинарного варианта (ФМ-2) широкое распространение получили ансамбли сигналов, построенные на основе матрицы Адамара (Hadamard). Матрицу Адамара размером 2^N можно построить следующим итерационным способом. Матрица размером 2×2 имеет вид

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Легко убедиться в том, что ее строки (или столбцы) ортогональны друг другу. Удвоение размера матрицы производится следующим образом:

$$H_{2n} = \begin{bmatrix} H_n & H_n \\ H_n & -H_n \end{bmatrix}.$$

Ортогональность строк (или столбцов) этой матрицы легко доказывается методом математической индукции.

Таким образом, с помощью матрицы Адамара можно получить ансамбли ортогональных бинарных сигналов. Функции Уолша, которые также широко используются в CDMA-системах, представляют собой те же самые строки матрицы Адамара, отличие лишь в способе упорядочения набора сигналов.

В направлении «вверх» чрезвычайно сложно обеспечить временную синхронизацию, поэтому используется *асинхронный* вариант CDMA (A-CDMA). Из-за отсутствия синхронизации сигналы разных пользователей перестают быть ортогональными. Соответственно, и приемник должен строиться по-другому.

Простейший вариант — *однопользовательский приемник* (Single-User Detector, SUD), в котором сигналы других пользователей трактуются просто

как шум. Если пользователей много, согласно центральной предельной теореме происходит нормализация статистических свойств этой *помехи множественного доступа* (Multiple Access Interference, MAI). При этом необходимо, чтобы сигналы всех пользователей приходили на базовую станцию с примерно одинаковым уровнем. Чтобы обеспечить это, в системах CDMA используется *управление мощностью* — по команде базовой станции мобильный терминал меняет мощность своего передатчика. Приемник в данном случае не имеет особенностей по сравнению со стандартным случаем (двухточечная линия связи без помех множественного доступа).

Отношение сигнал/помеха (по мощности) для такой системы с K пользователями и коэффициентом расширения спектра G примерно равно [3, 4.5]:

$$\text{SIR} \approx \frac{2G}{K-1}.$$

Строго говоря, оптимальным в данном случае будет *многопользовательский приемник* (MultiUser Detector, MUD), который осуществляет *совместную* оценку сигналов всех пользователей [3, 10.1]. Эта задача во многом аналогична приему сигнала при наличии МСИ (см. разд. 3), только в данном случае вместо помехи от соседних символов возникает помеха от других пользователей. Поэтому аналогичными оказываются и методы решения этих задач.

Оптимальный прием по методу МП может быть реализован согласно алгоритму Витерби. Состояния решетки в данном случае — это все возможные комбинации бит, переданных всеми пользователями. Таким образом, размер решетки экспоненциально увеличивается с ростом числа пользователей, поэтому практическая реализация данного метода возможна лишь при малом числе пользователей.

Прочие методы MUD строятся на базе фильтров и по смыслу своей работы аналогичны эквалайзерам, рассмотренным в разд. 3:

- *декоррелирующий приемник* (Decorrelating Detector) осуществляет инверсию матрицы взаимных корреляций между сигналами, преобразуя их к ортогональному набору. Это возможно, если сигналы линейно независимы. По смыслу это аналогично линейному эквалайзеру с форсированием нуля (ZFE);
- *приемник с минимумом среднеквадратической ошибки* (MMSE Detector) минимизирует не саму помеху множественного доступа, а суммарную погрешность воспроизведения выделяемого сигнала, т. е. сумму уровней MAI и шума. Это аналог линейного эквалайзера, рассчитываемого с учетом уровня шума;
- *приемник с последовательным вычитанием помех* (Successive Interference Cancellation Detector) сначала оценивает самый сильный сигнал, затем эта оценка вычитается, оценивается следующий сигнал и т. д. В определенной степени это аналогично эквалайзеру с ОСР.

Если в ортогональных системах число пользователей жестко ограничено, в асинхронных системах CDMA такого ограничения нет. Здесь могут использоваться и неортогональные сигналы. С ростом числа пользователей взаимные помехи будут плавно расти, а качество связи — плавно ухудшаться для всех пользователей.

Примером использования технологии CDMA являются системы мобильной связи третьего поколения (стандарты UMTS и cdma2000).

6.3. Случайный доступ

Под случайным доступом к среде передачи подразумевается такая организация многопользовательской системы, при которой отдельным пользователям не выделяются отдельные фрагменты частотно-временного ресурса. Вместо этого каждый пользователь в случайные моменты времени пытается передать сигнал. Доставка сообщения контролируется по поступлению подтверждения от приемной стороны по обратному каналу связи.

Так как пользователи осуществляют передачу случайным образом, возможно возникновение конфликтов (*коллизий*), при которых несколько пакетов перекрываются во времени. Если из-за этого пакет не будет принят, передающая сторона не дожидается подтверждения и через некоторое (случайное) время повторит попытку передачи.

Описанный простейший метод случайного (произвольного) доступа был разработан в Гавайском университете в 1971 г. и получил название ALOHA.

Определим возможную загрузку среды передачи при использовании ALOHA, считая, что суммарная интенсивность поступления пакетов от всех пользователей в среднем равна λ пакетов в единицу времени, длительности пакетов одинаковы и равны τ , а статистика их поступления описывается законом Пуассона, т. е. вероятность того, что за время Δt от всех пользователей в сумме поступит k пакетов:

$$P(k, \Delta t) = \frac{(\lambda \Delta t)^k}{k!} e^{-\lambda \Delta t}.$$

Считаем, что пакеты при отсутствии коллизий передаются без ошибок, а при коллизии пакет теряется. Коллизия будет отсутствовать, если при передаче пакета не возникло «конкуренции» на протяжении интервала длительностью 2τ . Таким образом, вероятность отсутствия коллизии (а следовательно, и успешной доставки пакета):

$$P_{\text{дост}} = P(0, 2\tau) = e^{-2\lambda\tau}.$$

Частота успешной доставки пакетов будет равна $\lambda P_{\text{дост}}$, а доля времени, занятого успешно доставленными пакетами:

$$\rho = \lambda \tau P_{\text{дост}} = \lambda \tau e^{-2\lambda\tau}. \quad (6.1)$$

Зависимость (6.1) показана на рис. 6.2 сплошной линией, она имеет максимум при $\lambda\tau = 0,5$, равный $1/(2e) \approx 0,18$. При интенсивности поступления пакетов ниже оптимальной коллизии будут происходить редко, но канал будет часто простаивать. Если интенсивность поступления пакетов больше оптимальной, часто происходящие коллизии снижают эффективность использования канала. Но даже в оптимальном случае успешно доставленными пакетами будет занято лишь 18 % времени.

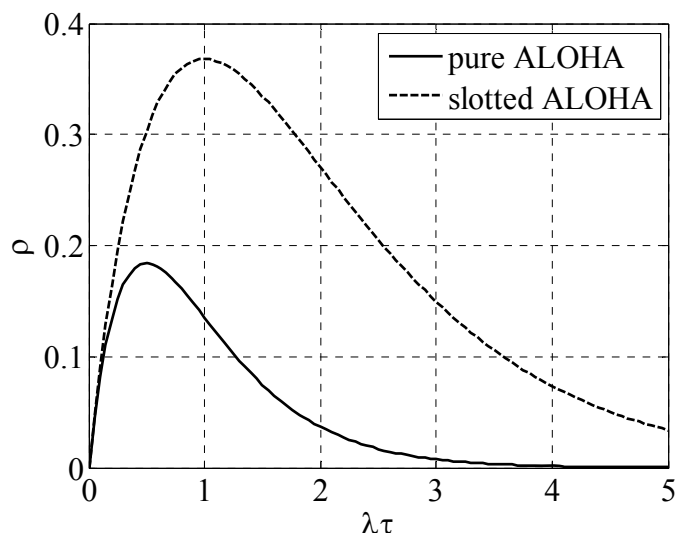


Рис. 6.2. Эффективность использования среды передачи при случайном доступе (ALOHA)

Описанный вариант называется *чистой* ALOHA (pure ALOHA). Попытка улучшить эффективность использования среды передачи без заметного усложнения алгоритма привела к разработке версии ALOHA *с выделением временных интервалов* (slotted ALOHA). Идея состоит в том, что временная ось делится на промежутки (*слоты*) длительностью τ и передача всех пакетов, поступивших на протяжении одного слота, задерживается до начала следующего слота. Для этого, естественно, в системе связи необходима временная синхронизация между всеми пользователями.

За счет привязки моментов передачи пакетов к единой временной сетке оказывается невозможным частичное перекрытие пакетов и «конфликтный» промежуток сокращается до τ . Выкладки, аналогичные приведенным ранее, дают в данном случае

$$\rho = \lambda\tau e^{-\lambda\tau}. \quad (6.2)$$

Зависимость (6.2) показана на рис. 6.2 штриховой линией, она имеет максимум при $\lambda\tau = 1$, равный $1/e \approx 0,37$. Таким образом, в данном случае удастся достичь вдвое большей эффективности использования канала, чем при «чистой» ALOHA.

7. СИНХРОНИЗАЦИЯ

Ранее при рассмотрении процессов демодуляции предполагалось, что отсчеты сигнала берутся точно в нужные моменты времени, а получаемые в результате точки на комплексной плоскости правильно расположены относительно координатных осей. В реальных системах это обеспечивается за счет работы систем синхронизации.

7.1. Задачи синхронизации

К основным видам синхронизации, работающим на физическом уровне системы связи, относятся:

- частотная синхронизация;
- фазовая синхронизация;
- временная (тактовая) синхронизация.

Необходимость в отдельной частотной синхронизации возникает в тех случаях, когда в течение символического интервала фазовый сдвиг заметно меняется. Если же его на таком промежутке можно считать постоянным, то задача частотной синхронизации может решаться с помощью системы фазовой синхронизации при соответствующем числе интеграторов в петле автоподстройки.

Задачи синхронизации в системах цифровой связи тесно связаны с вопросами, рассматриваемыми в следующих дисциплинах:

- **Статистическая радиотехника [6]:** задачи синхронизации по сути дела представляют собой задачи оценивания параметров сигнала. Особенность, присущая связным системам, как и при рассмотрении процесса демодуляции как задачи различения сигналов, состоит в том, что сигнальные посылки непрерывно следуют друг за другом.
- **Системы радиоавтоматики [7]:** задача коррекции частотных, фазовых и временных ошибок часто решается с помощью замкнутых следящих систем, подробный анализ работы которых производится в указанной дисциплине. Особенностью работы следящих систем в связных приложениях является то, что сигнал ошибки в большинстве случаев содержит в себе случайную составляющую, обусловленную случайным характером передаваемых данных и затрудняющую точный анализ процессов синхронизации.

В принципе, для синхронизации системы связи может использоваться специальный синхронизирующий сигнал, генерируемый специально для облегчения данного процесса. Однако данный подход применяется на практике редко, так как на передачу такого сигнала тратится дополнительная мощность. Как правило, синхронизация выполняется непосредственно по информационному сигналу. Одной из функций скремблирования, рассмотренного

ранее, как раз и является облегчение синхронизации (прежде всего временной) благодаря рандомизации данных и практическому исключению возможности генерации постоянного (или гармонического) сигнала в течение длительного времени.

7.2. Ухудшение помехоустойчивости из-за ошибок синхронизации

Рассмотрим причины ухудшения помехоустойчивости систем цифровой связи при наличии ошибок фазовой и временной синхронизации.

Ошибка фазовой синхронизации. Проще всего показать влияние этой ошибки на примере сигнала с модуляцией ФМ-2. На рис. 7.1 показано положение точек созвездия ФМ-2 при отклонении фазы принимаемого сигнала на угол φ .

Из рисунка видно, что расстояние до границы между областями принятия решения о нулевом/единичном значении переданного бита при наличии фазовой ошибки уменьшается пропорционально ее косинусу. Таким образом, в формуле для вероятности битовой ошибки необходимо умножить аргумент Q -функции на $\cos \varphi$:

$$P_b(\varphi) = Q\left(\sqrt{\frac{2E_b}{N_0}} \cos \varphi\right). \quad (7.1)$$

Если фазовая ошибка является случайной величиной с плотностью вероятности $p_\varphi(\varphi)$, нужно усреднить (7.1) в соответствии с вероятностными свойствами φ :

$$\overline{P_b(\varphi)} = \int_{-\infty}^{\infty} P_b(\varphi) p_\varphi(\varphi) d\varphi = \int_{-\infty}^{\infty} Q\left(\sqrt{\frac{2E_b}{N_0}} \cos \varphi\right) p_\varphi(\varphi) d\varphi.$$

Для гауссова распределения ошибки это выражение принимает вид

$$\overline{P_b(\varphi)} = \int_{-\infty}^{\infty} Q\left(\sqrt{\frac{2E_b}{N_0}} \cos \varphi\right) \frac{1}{\sqrt{2\pi}\sigma_\varphi} e^{-\frac{\varphi^2}{2\sigma_\varphi^2}} d\varphi,$$

где σ_φ — среднеквадратическое значение фазовой ошибки.

Этот интеграл может быть вычислен только численно, результаты для нескольких значений σ_φ показаны на рис. 7.2.

Главной особенностью полученных результатов в данном случае является то, что вероятность ошибки при возрастании отношения сигнал/шум стремится не к нулю, а к некоторому ненулевому значению.

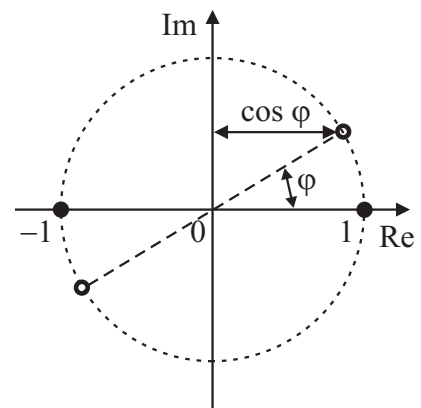


Рис. 7.1. Фазовая ошибка при приеме сигнала с ФМ-2

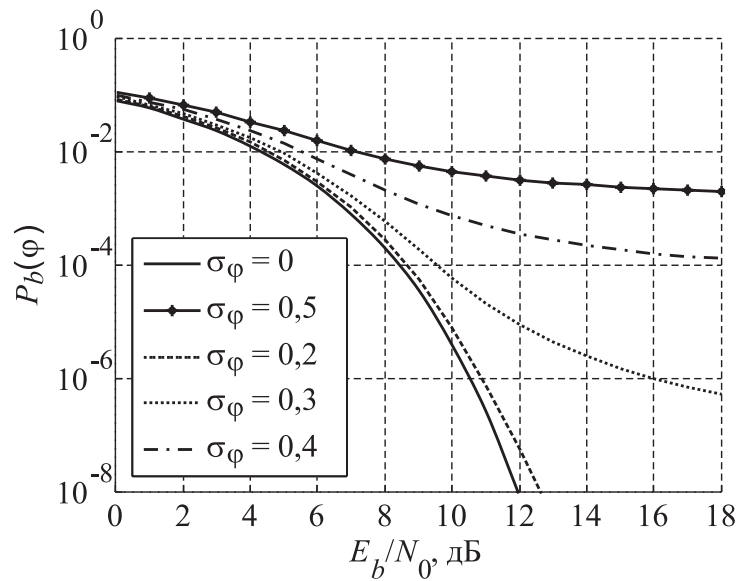


Рис. 7.2. Помехоустойчивость системы связи с ФМ-2 при наличии гауссовой фазовой ошибки

Ошибка временной синхронизации. Данный вид ошибки проще всего рассмотреть на примере низкочастотной бинарной амплитудно-импульсной модуляции с прямоугольными посылками. Те же количественные соотношения будут справедливы для модуляции ФМ-2, а общий смысл рассуждений сохраняется и при других видах линейной модуляции.

На рис. 7.3 показано взаимное временное положение сигнальных посылок и опорного сигнала коррелятора приемника (заштрихованный прямоугольник) при наличии временной ошибки на τ .

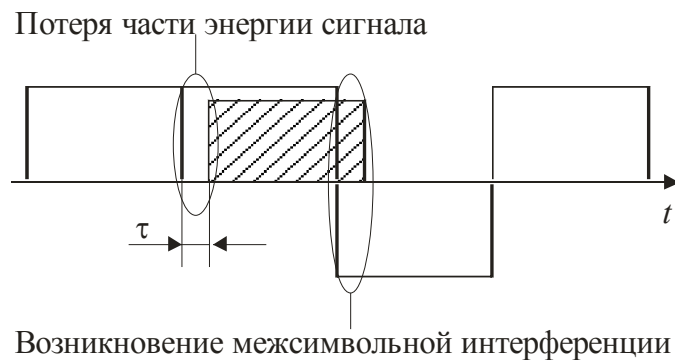


Рис. 7.3. Рассогласование по времени между принятой посылкой и опорным сигналом коррелятора

Из рисунка видно, что влияние временной ошибки проявляется двояко:

- теряется часть энергии полезного сигнала;
- возникает межсимвольная интерференция из-за наложения опорного сигнала на соседнюю сигнальную посылку.

Математически происходящие при этом процессы можно описать следующим образом. Сигнал после согласованного фильтра имеет вид (см. 2.1.7 и формулу (2.13))

$$y(t) = \sum_{k=-\infty}^{\infty} a_k B_g(t - kT),$$

где $B_g(t)$ — корреляционная функция сигнальной посылки $g(t)$.

В идеальном случае необходимо брать отсчеты этого сигнала в моменты времени $t_n = nT$, когда достигаются максимумы отклика согласованного фильтра на отдельные символы и отсутствует межсимвольная интерференция (если, разумеется, сигнальная посылка удовлетворяет соответствующим условиям, рассмотренным в 2.1.7). Пусть имеется фиксированная ошибка временной синхронизации и отсчеты берутся в моменты $t_n = nT + \tau$. В результате получим следующий сигнал:

$$y(t_n) = \sum_{k=-\infty}^{\infty} a_k B_g((n-k)T + \tau). \quad (7.2)$$

Таким образом, вклад от ожидаемого в данный момент символа a_n уменьшается с коэффициентом, равным $B_g(\tau)/B_g(0)$; кроме того, появляется МСИ от соседних символов a_k , $k \neq n$. Соответствующая иллюстрация для случая прямоугольной сигнальной посылки приведена на рис. 7.4.

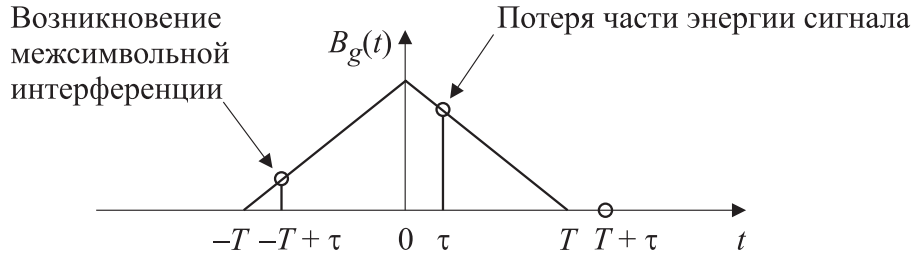


Рис. 7.4. Вклады от различных символов при неточной временной синхронизации

Число символов, дающих вклад в возникающую МСИ, определяется длительностью сигнальной посылки. Для прямоугольной посылки длительностью T такой символ только один, и (7.2) можно записать в виде

$$y(t_n) = a_n B_g(\tau) + a_{n+1} B_g(-T + \tau) = a_n E_g \left(1 - \frac{|\tau|}{T} \right) + a_{n+1} E_g \frac{|\tau|}{T}.$$

Здесь предполагалось, что $0 \leq \tau \leq T$; для отрицательных τ , лежащих в диапазоне $-T \leq \tau < 0$, вместо следующего символа a_{n+1} вклад в МСИ будет давать (с таким же коэффициентом) предыдущий символ a_{n-1} .

Помеха от МСИ является негауссовой, и ее влияние на вероятность ошибки в рассматриваемом примере (ФМ-2 и прямоугольная сигнальная по-

сылка) можно описать следующим образом. С вероятностью 1/2 соседний символ, вызывающий МСИ, принимает то же значение, что и символ a_n , в результате принятый сигнал

$$y(t_n) = a_n (B_g(\tau) + B_g(-T + \tau)).$$

С такой же вероятностью 1/2 соседний символ будет принимать противоположное значение $-a_n$, так что отсчет принимаемого сигнала

$$y(t_n) = a_n (B_g(\tau) - B_g(-T + \tau)).$$

На уровень шума временная ошибка никак не влияет, поэтому вероятность ошибки будет изменяться только из-за того, что уровень полезного сигнала в двух рассмотренных ситуациях оказывается отличным от ожидаемого в идеальном случае значения $a_n B_g(0) = a_n E_g$. Умножая аргумент Q -функции, описывающей вероятность ошибки при модуляции ФМ-2, на соответствующий коэффициент, и усредняя 2 рассмотренных варианта, получаем следующую формулу для вероятности ошибки:

$$\overline{P_b(\tau)} = \frac{1}{2} Q \left(\sqrt{\frac{2E_b}{N_0}} \frac{B_g(\tau) + B_g(-T + \tau)}{E_g} \right) + \frac{1}{2} Q \left(\sqrt{\frac{2E_b}{N_0}} \frac{B_g(\tau) - B_g(-T + \tau)}{E_g} \right).$$

Конкретизируя эту формулу для случая прямоугольной сигнальной посылки, получаем

$$\overline{P_b(\tau)} = \frac{1}{2} Q \left(\sqrt{\frac{2E_b}{N_0}} \right) + \frac{1}{2} Q \left(\sqrt{\frac{2E_b}{N_0}} \left(1 - 2 \frac{|\tau|}{T} \right) \right), \quad |\tau| \leq T. \quad (7.3)$$

На рис. 7.5 показаны графики помехоустойчивости, построенные по формуле (7.3) для нескольких значений отношения τ/T .

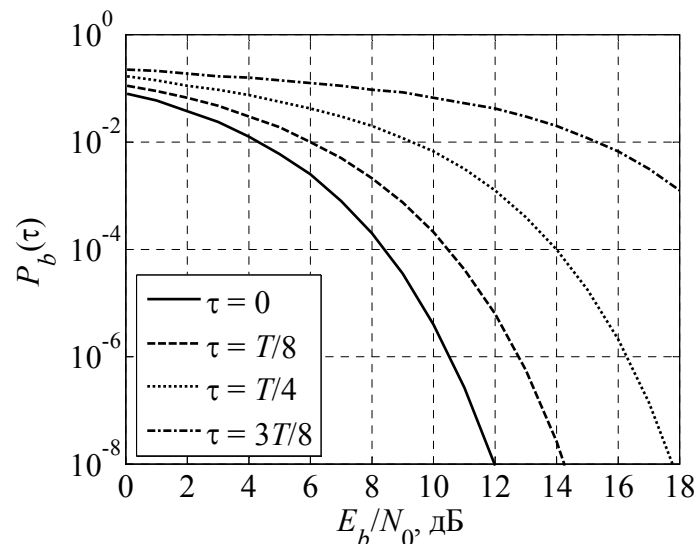


Рис. 7.5. Помехоустойчивость системы связи с ФМ-2 при наличии фиксированной временной ошибки

Из графиков видно, что уже при временной ошибке в $1/8$ длительности символа энергетические потери составляют около 2 дБ. С увеличением ошибки потери стремительно возрастают, а при $\tau = T/2$ вероятность ошибки перестает стремиться к нулю — согласно (7.3), в данном случае получается

$$\overline{P_b(T/2)} = \frac{1}{2} Q\left(\sqrt{\frac{2E_b}{N_0}}\right) + \frac{1}{4}.$$

Если ошибка временной синхронизации является случайной величиной с плотностью вероятности $p_\tau(\tau)$, усреднение вероятности ошибки производится так же, как в случае фазовой ошибки:

$$\overline{P_b(\tau)} = \int_{-\infty}^{\infty} P_b(\tau) p_\tau(\tau) d\tau.$$

На рис. 7.6 показано, как влияет наличие гауссовой временной ошибки на помехоустойчивость системы связи с ФМ-2. Параметр σ_τ на графиках — это среднеквадратическое значение временной ошибки, нормированное к длительности символа.

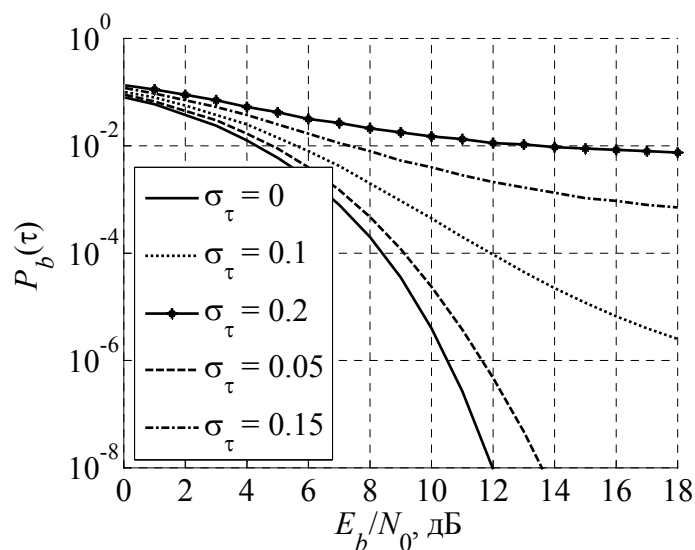


Рис. 7.6. Помехоустойчивость системы связи с ФМ-2 при наличии гауссовой временной ошибки

В данном случае, как и при наличии случайной фазовой ошибки (см. рис. 7.2), с ростом отношения сигнал/шум вероятность ошибки стремится не к нулю, а к некоторому ненулевому значению.

7.3. Классификация алгоритмов синхронизации

Алгоритмы синхронизации можно классифицировать по различным признакам:

- **Алгоритмы с использованием и без использования данных** — в зависимости от того, считаются ли известными значения передаваемых символов. Алгоритмы с использованием данных могут быть реализованы в двух вариантах:
 1. Данные действительно *известны*, если синхронизация производится по специальной *пreamбуле*, которая генерируется передатчиком по известному на приемной стороне закону.
 2. Вместо истинных значений данных могут использоваться их *оценки*, получаемые после демодуляции. В этом случае реализуется синхронизация с ОСР.

В алгоритмах без использования данных (их еще называют *слепыми* (blind) алгоритмами) передаваемые символы считаются случайными и никаких предположений об их значениях не делается.

- **Алгоритмы разомкнутые и замкнутые.** *Разомкнутые* (англ. термин — feedforward) алгоритмы оценивают параметр, затем результат оценивания используется для коррекции соответствующей ошибки. Обратная связь в таких системах отсутствует (рис. 7.7, а). *Замкнутые* (англ. термин — feedback) алгоритмы (рис. 7.7, б) представляют собой следящие системы, в которых принимаемый сигнал сначала подвергается коррекции нужного параметра, а затем из скорректированного сигнала вырабатывается *сигнал ошибки*. В качестве этого сигнала, в принципе, может выступать любой сигнал, удовлетворяющий следующим требованиям: при отсутствии ошибки его математическое ожидание должно быть нулевым, при этом производная зависимости математического ожидания от значения ошибки должна быть отлична от нуля.

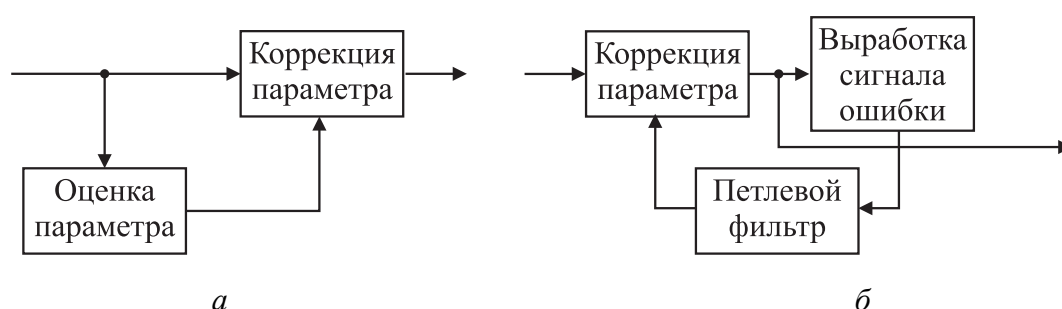


Рис. 7.7. Разомкнутая (а) и замкнутая (б) системы синхронизации

Зависимость математического ожидания сигнала ошибки от значения ошибки благодаря форме соответствующего графика называется *S-кривой*.

Далее выработанный сигнал ошибки поступает на петлевой фильтр (фильтр нижних частот), который его сглаживает. Выходной сигнал фильтра представляет собой оценку искомого параметра. Она поступает на блок коррекции параметра, замыкая петлю обратной связи.

Для отслеживания медленных изменений параметра в процессе работы, как правило, используются замкнутые алгоритмы. Разомкнутые алгоритмы обычно применяются для быстрой первоначальной оценки параметров.

- **Алгоритмы раздельной и совместной синхронизации.** Алгоритмы *раздельной* синхронизации работают с каким-то одним параметром, полагая остальные известными либо случайными. Алгоритмы *совместной* синхронизации совместно оценивают несколько параметров одновременно. Так, возможно совместное оценивание временной и фазовой ошибок.
- **Алгоритмы, предполагающие прочие параметры сигнала известными либо случайными.** В зависимости от того, в какой последовательности реализуются процедуры синхронизации, необходимо использовать соответствующие предположения о характере прочих (отличных от синхронизируемого на данном этапе) параметров сигнала. Так, если сначала реализуется временная синхронизация, а затем фазовая (эта последовательность часто используется на практике), то алгоритм временной синхронизации должен считать начальную фазу сигнала случайной, а алгоритм фазовой синхронизации может считать временное положение символов уже известным.
- **Алгоритмы, основанные на принципе максимального правдоподобия, и эвристические.** Вывод формул, по которым оценивается искомый параметр в разомкнутых системах либо формируется сигнал ошибки в замкнутых системах, может основываться на максимизации функции правдоподобия (разумеется, для оценки параметра существует не только метод МП, но именно он в большинстве случаев используется на практике), а может базироваться на каких-либо эвристических соображениях. Алгоритмы, выведенные на основе МП-подхода, в большинстве случаев не являются МП-методами вследствие сделанных в процессе их вывода приближений и допущений. Поэтому эвристические методы иногда могут обеспечивать лучшее качество работы, чем методы, полученные на основе МП-подхода.

7.4. Фазовая синхронизация

Рассмотрим несколько алгоритмов фазовой синхронизации для сигналов с линейной модуляцией, полученных на основе МП-подхода — с использованием и без использования данных, замкнутые и разомкнутые. Во всех случаях рассматривается сигнал после согласованного фильтра, имеющий один отсчет на символ (т. е. считается, что временная синхронизация уже выполнена). При наличии фазовой ошибки φ этот сигнал будет иметь вид

$$\dot{r}_k = \dot{a}_k e^{j\varphi} + \dot{n}_k,$$

где a_k — передаваемые символы сигнального созвездия, n_k — отсчеты АБГШ.

«Чистый» сигнал, не имеющий шума и фазовой ошибки, можно записать как

$$\dot{s}_k = \dot{a}_k .$$

7.4.1. Разомкнутый алгоритм с использованием данных

Пусть наблюдаемый интервал имеет длину K символов. Тогда совместная плотность вероятности отсчетов на выходе согласованного фильтра (при условии, что передавалась последовательность символов $\{a_k\}$ и фазовый сдвиг был равен φ)

$$p(\{\dot{r}_k\} | \{\dot{a}_k\}, \varphi) = \prod_{k=1}^K \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|r_k - a_k e^{j\varphi}|^2}{2\sigma^2}\right), \quad (7.4)$$

где σ^2 — половина дисперсии комплексного шума на выходе фильтра. МП-оценка параметра φ должна максимизировать эту ПВ, или, что то же самое, максимизировать ее логарифм:

$$\ln p(\{\dot{r}_k\} | \{\dot{a}_k\}, \varphi) = -\sum_{k=1}^K \frac{|r_k - a_k e^{j\varphi}|^2}{2\sigma^2} - K \ln(2\pi\sigma^2) \rightarrow \max_{\varphi} .$$

Отсюда видно, что для поиска МП-оценки необходимо минимизировать квадрат евклидова расстояния между принятой реализацией и «чистым» сигналом, в который внесены соответствующие фазовые искажения:

$$\sum_{k=1}^K |\dot{r}_k - \dot{a}_k e^{j\varphi}|^2 \rightarrow \min_{\varphi} .$$

Раскрытие квадрата модуля в левой части дает следующее:

$$\sum_{k=1}^K |\dot{r}_k|^2 + \sum_{k=1}^K |\dot{a}_k|^2 - 2 \operatorname{Re} \left(e^{-j\varphi} \sum_{k=1}^K \dot{r}_k \dot{a}_k^* \right) \rightarrow \min_{\varphi} .$$

Первые 2 слагаемых не зависят от параметра φ , поэтому поиск решения сводится к максимизации корреляционной суммы:

$$\operatorname{Re} \left(e^{-j\varphi} \sum_{k=1}^K \dot{r}_k \dot{a}_k^* \right) \rightarrow \max_{\varphi} . \quad (7.5)$$

Вещественная часть может быть записана следующим образом:

$$\left| \sum_{k=1}^K \dot{r}_k \dot{a}_k^* \right| \cos \left(\arg \left(\sum_{k=1}^K \dot{r}_k \dot{a}_k^* \right) - \varphi \right) \rightarrow \max_{\varphi} . \quad (7.6)$$

Модуль комплексной корреляции не зависит от фазового сдвига, поэтому остается следующий критерий:

$$\cos \left(\arg \left(\sum_{k=1}^K \dot{r}_k \dot{a}_k^* \right) - \varphi \right) \rightarrow \max_{\varphi}.$$

Максимальное значение косинуса равно единице, и достигается оно при нулевом аргументе. Отсюда получаем формулу для оценки фазового сдвига:

$$\hat{\varphi} = \arg \left(\sum_{k=1}^K \dot{r}_k \dot{a}_k^* \right). \quad (7.7)$$

Данная статистика позволяет оценить фазовый сдвиг при известных данных a_k независимо от использованного вида линейной модуляции.

7.4.2. Замкнутый алгоритм с использованием данных

Для работы следящей системы необходимо сформировать сигнал ошибки, равный нулю при истинном значении параметра. Для получения такого сигнала продифференцируем левую часть выражения (7.5), которое в разомкнутом варианте системы подвергалось максимизации. Получим следующее:

$$\begin{aligned} \varepsilon &= \frac{d}{d\varphi} \left(\operatorname{Re} \left(e^{-j\varphi} \sum_{k=1}^K \dot{r}_k \dot{a}_k^* \right) \right) = \operatorname{Re} \left(-j e^{-j\varphi} \sum_{k=1}^K \dot{r}_k \dot{a}_k^* \right) = \\ &= \operatorname{Im} \left(e^{-j\varphi} \sum_{k=1}^K \dot{r}_k \dot{a}_k^* \right) = \operatorname{Im} \left(\sum_{k=1}^K (\dot{r}_k e^{-j\varphi}) \dot{a}_k^* \right). \end{aligned} \quad (7.8)$$

В этом выражении $\dot{r}_k e^{-j\varphi}$ — не что иное, как входной сигнал после коррекции фазовой ошибки. Обозначим его как \dot{r}'_k и запишем результат дифференцирования (7.8) в виде

$$\varepsilon = \operatorname{Im} \left(\sum_{k=1}^K \dot{r}'_k \dot{a}_k^* \right) = \sum_{k=1}^K \operatorname{Im}(\dot{r}'_k \dot{a}_k^*).$$

Правая часть этого выражения предполагает суммирование (т. е. усреднение) по времени. В замкнутых системах это усреднение возлагается на петлевой фильтр, поэтому в качестве сигнала ошибки используется одиночное слагаемое полученной формулы:

$$e_k = \operatorname{Im}(\dot{r}'_k \dot{a}_k^*).$$

Структурная схема получившейся системы фазовой синхронизации показана на рис. 7.8.

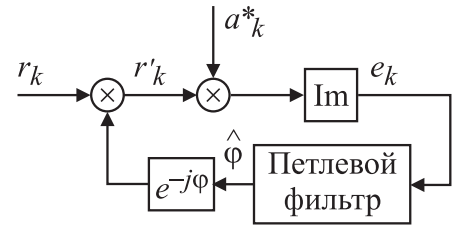


Рис. 7.8. Замкнутая система фазовой синхронизации с использованием данных

Как и в случае разомкнутого варианта (7.7), эта замкнутая система не зависит от используемого вида линейной модуляции.

7.4.3. Разомкнутый алгоритм без использования данных

Чтобы перейти к алгоритму, не использующему символы a_k , в (7.4) необходимо выполнить усреднение по возможным символам созвездия $a_k \in \{C_m\}$, $m = 1, 2, \dots, M$:

$$p(\{\dot{r}_k\} | \varphi) = \prod_{k=1}^K \frac{1}{2\pi\sigma^2} \left(\frac{1}{M} \sum_{m=1}^M \exp\left(-\frac{|r_k - C_m e^{j\varphi}|^2}{2\sigma^2}\right) \right).$$

Дальнейшее рассмотрение будет произведено для модуляции ФМ-2, когда $M = 2$ и $C_m = \{+1, -1\}$. В этом случае

$$p(\{\dot{r}_k\} | \varphi) = \prod_{k=1}^K \frac{1}{2\pi\sigma^2} \left(\frac{1}{2} \exp\left(-\frac{|r_k - e^{j\varphi}|^2}{2\sigma^2}\right) + \frac{1}{2} \exp\left(-\frac{|r_k + e^{j\varphi}|^2}{2\sigma^2}\right) \right).$$

Логарифмирование этого выражения дает

$$\ln p(\{\dot{r}_k\} | \varphi) = \sum_{k=1}^K \ln \left(\exp\left(-\frac{|r_k - e^{j\varphi}|^2}{2\sigma^2}\right) + \exp\left(-\frac{|r_k + e^{j\varphi}|^2}{2\sigma^2}\right) \right) - K \ln(4\pi\sigma^2).$$

Второе слагаемое не зависит от фазы, поэтому для поиска оценки необходимо максимизировать первое слагаемое:

$$\sum_{k=1}^K \ln \left(\exp\left(-\frac{|r_k - e^{j\varphi}|^2}{2\sigma^2}\right) + \exp\left(-\frac{|r_k + e^{j\varphi}|^2}{2\sigma^2}\right) \right) \rightarrow \max_{\varphi}. \quad (7.9)$$

В общем случае аналитического решения не получить. Однако можно приближенно рассмотреть асимптотический случай малого отношения сигнал/шум. Для начала переформулируем (7.9), вынеся общие множители из экспонент:

$$\begin{aligned} & \sum_{k=1}^K \ln \left(\exp\left(-\frac{|r_k|^2 + 1}{2\sigma^2}\right) \left(\exp\left(\frac{\operatorname{Re}(r_k e^{-j\varphi})}{2\sigma^2}\right) + \exp\left(-\frac{\operatorname{Re}(r_k e^{-j\varphi})}{2\sigma^2}\right) \right) \right) = \\ & = \sum_{k=1}^K \ln \left(\exp\left(\frac{\operatorname{Re}(r_k e^{-j\varphi})}{2\sigma^2}\right) + \exp\left(-\frac{\operatorname{Re}(r_k e^{-j\varphi})}{2\sigma^2}\right) \right) - \sum_{k=1}^K \frac{|r_k|^2 + 1}{2\sigma^2} \rightarrow \max_{\varphi}. \end{aligned}$$

Далее устраняем слагаемое, не зависящее от φ :

$$\sum_{k=1}^K \ln \left(\exp\left(\frac{\operatorname{Re}(r_k e^{-j\varphi})}{2\sigma^2}\right) + \exp\left(-\frac{\operatorname{Re}(r_k e^{-j\varphi})}{2\sigma^2}\right) \right) \rightarrow \max_{\varphi}.$$

Теперь воспользуемся предположением о малости отношения сигнал/шум. При сильном шуме показатели экспонент близки к нулю и можно квадратично аппроксимировать экспоненты и логарифм усеченными рядами Тейлора:

$$e^x \approx 1 + x + \frac{x^2}{2}, \quad e^x + e^{-x} \approx 2 + x^2,$$

$$\ln(2 + x^2) = \ln 2 + \ln(1 + x^2/2) \approx \ln 2 + x^2/2 \quad \text{при } |x| \ll 1.$$

Использование этих приближений дает:

$$K \ln 2 + \sum_{k=1}^K \frac{1}{2} \left(\frac{\operatorname{Re}(r_k e^{-j\varphi})}{2\sigma^2} \right)^2 \rightarrow \max_{\varphi}.$$

Отбрасывая незначимые слагаемые и множители, получаем

$$\sum_{k=1}^K \operatorname{Re}^2(r_k e^{-j\varphi}) \rightarrow \max_{\varphi}. \quad (7.10)$$

С левой частью (7.10) можно выполнить следующие преобразования:

$$\begin{aligned} \sum_{k=1}^K \operatorname{Re}^2(r_k e^{-j\varphi}) &= \sum_{k=1}^K |r_k|^2 \cos^2(\arg r_k - \varphi) = \\ &= \frac{1}{2} \sum_{k=1}^K |r_k|^2 + \frac{1}{2} \sum_{k=1}^K |r_k|^2 \cos(2 \arg r_k - 2\varphi) = \\ &= \frac{1}{2} \sum_{k=1}^K |r_k|^2 + \frac{1}{2} \operatorname{Re} \left(\sum_{k=1}^K r_k^2 e^{-j2\varphi} \right) \rightarrow \max_{\varphi}. \end{aligned} \quad (7.11)$$

Опять отбрасывая несущественные слагаемые и множители, получаем

$$\operatorname{Re} \left(\sum_{k=1}^K r_k^2 e^{-j2\varphi} \right) = \left| \sum_{k=1}^K r_k^2 \right| \cos \left(\arg \left(\sum_{k=1}^K r_k^2 \right) - 2\varphi \right) \rightarrow \max_{\varphi}.$$

Отсюда, по аналогии со случаем известных данных (7.6), получаем

$$\hat{\varphi} = \frac{1}{2} \arg \left(\sum_{k=1}^K r_k^2 \right). \quad (7.12)$$

Этот вариант алгоритма, как видим, не использует информацию о переданных символах \dot{a}_k .

Принципиальным отличием от случая известных данных является тот факт, что, согласно (7.12), $|\hat{\varphi}| \leq \pi/2$, т. е. получаемая оценка фазы лежит лишь в пределах половины окружности. Дело в том, что при неизвестных данных невозможно отличить «истинное» положение сигнального созвездия ФМ-2 от

повернутого на 180° варианта. Аналогичные проблемы неопределенности возникают для любых алгоритмов фазовой синхронизации без использования данных, если сигнальное созвездие обладает угловой симметрией (а таковы все созвездия, используемые на практике). Для разрешения неопределенности фазовой синхронизации применяется дифференциальное кодирование данных, когда информация заключена не в абсолютной фазе сигнала, а в ее изменении от символа к символу.

Идея формулы (7.12) состоит в «снятии модуляции» с сигнала ФМ-2 возведением его в квадрат. Эта идея может быть очевидным образом расширена на случай M -позиционной ФМ:

$$\hat{\phi} = \frac{1}{M} \arg \left(\sum_{k=1}^K \dot{r}_k^M \right). \quad (7.13)$$

Формулу (7.13) можно использовать и для КАМ с квадратными созвездиями, задав $M = 4$. Качество получаемой оценки будет хуже, чем для ФМ-сигналов, так как полного снятия модуляции в данном случае не происходит — возведенный в четвертую степень сигнал содержит случайную составляющую, связанную с передаваемыми символами. Из-за этого оценка (7.13) для КАМ-сигналов будет иметь *собственный шум* — ее дисперсия с ростом отношения сигнал/шум будет стремиться не к нулю, а к некоторому ненулевому значению.

7.4.4. Замкнутый алгоритм без использования данных

Вывод формулы для сигнала ошибки аналогичен случаю известных данных (см. 7.4.2). Дифференцируя (7.11), получаем

$$\begin{aligned} \varepsilon &= \frac{d}{d\phi} \left(\frac{1}{2} \sum_{k=1}^K |\dot{r}_k|^2 + \frac{1}{2} \operatorname{Re} \left(\sum_{k=1}^K \dot{r}_k^2 e^{-j2\phi} \right) \right) = \\ &= \frac{1}{2} \operatorname{Re} \left(-2j \sum_{k=1}^K \dot{r}_k^2 e^{-j2\phi} \right) = \operatorname{Im} \left(\sum_{k=1}^K \dot{r}_k^2 e^{-j2\phi} \right) = \operatorname{Im} \left(\sum_{k=1}^K \left(\dot{r}_k e^{-j\phi} \right)^2 \right). \end{aligned}$$

Оставляя в этом выражении одиночный член суммы, имеем

$$e_k = \operatorname{Im} \left(\left(\dot{r}_k e^{-j\phi} \right)^2 \right) = \operatorname{Im} \left(\left(\dot{r}'_k \right)^2 \right) = 2 \operatorname{Re}(\dot{r}'_k) \operatorname{Im}(\dot{r}'_k).$$

Отбрасывая несущественный постоянный множитель, окончательно получаем

$$e_k = \operatorname{Re}(\dot{r}'_k) \operatorname{Im}(\dot{r}'_k).$$

Структурная схема получившейся системы слепой фазовой синхронизации показана на рис. 7.9.

Похожим образом могут быть выведены формулы для сигнала ошибки и в случае иной, чем при ФМ-2, угловой симметрии сигнального созвездия.

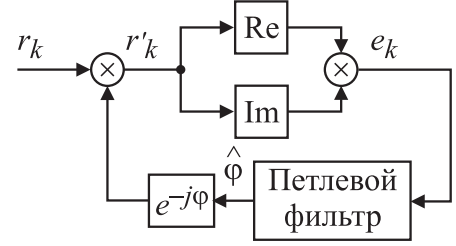


Рис. 7.9. Замкнутая система фазовой синхронизации без использования данных для сигнала с модуляцией ФМ-2

7.5. Временная синхронизация

Рассматриваемые далее алгоритмы временной синхронизации предназначены для сигналов с линейной модуляцией. При этом частота и фаза сигнала считаются известными.

7.5.1. Разомкнутый алгоритм с использованием данных

В данном случае фактически ищется временное положение известного сигнала. Согласно теории оценки параметров, если сигнал искажен АБГШ-каналом, для этого нужно найти такой временной сдвиг τ , при котором оказывается максимальной вещественная часть взаимной корреляционной функции (ВКФ) между принятым сигналом $s(t)$ и известным на приемной стороне образцом $s_0(t)$.

Запишем ВКФ, которую в данном случае необходимо максимизировать, с учетом того, что сигнал представляет собой сигнал с цифровой линейной модуляцией вида (2.1):

$$\begin{aligned} B_{ss_0}(\tau) &= \int_{-\infty}^{\infty} s(t) s_0^*(t - \tau) dt = \int_{-\infty}^{\infty} s(t) \sum_k a_k^* g^*(t - \tau - kT) dt = \\ &= \sum_k a_k^* \int_{-\infty}^{\infty} s(t) g^*(t - \tau - kT) dt. \end{aligned}$$

Входящий в полученное выражение интеграл представляет собой значение сигнала $y(t)$ на выходе фильтра, согласованного с сигнальной посылкой, вычисленное в момент времени $kT + \tau$, поэтому ВКФ можно записать так:

$$B_{ss_0}(\tau) = \sum_k a_k^* y(kT + \tau).$$

В результате для оценки временного сдвига сигнала при известных данных получаем следующее правило:

$$\operatorname{Re}(B_{ss_0}(\tau)) = \operatorname{Re}\left(\sum_k a_k^* y(kT + \tau)\right) \rightarrow \max_{\tau}, \quad (7.14)$$

где $y(t)$ — сигнал на выходе согласованного фильтра приемника.

7.5.2. Замкнутый алгоритм с использованием данных

Дифференцируя (7.14) по τ , получаем сигнал, пригодный для использования в качестве сигнала ошибки в замкнутой системе временной синхронизации:

$$\varepsilon = \frac{d}{d\tau} \operatorname{Re} \left(\sum_k \dot{a}_k^* y(kT + \tau) \right) = \operatorname{Re} \left(\sum_k \dot{a}_k^* y'(kT + \tau) \right), \quad (7.15)$$

где $y'(t)$ — производная по времени от сигнала на выходе СФ. Можно представить формирование этого сигнала как пропускание входного сигнала через дифференцирующий СФ, ИХ которого является производной по времени от ИХ согласованного фильтра. Так как ИХ СФ есть перевернутая вдоль временной оси сигнальная посылка, $h(t) = g(-t)$, то производная от нее имеет вид $h'(t) = -g'(-t)$. Таким образом, сигнал $y'(t)$ можно получить, пропуская входной сигнал через фильтр с указанной ИХ.

Как обычно делается при формировании сигнала ошибки для замкнутых систем синхронизации, усреднение по времени возлагается на петлевой фильтр, поэтому в качестве сигнала ошибки e_k используется одиночное слагаемое (7.15):

$$e_k = \operatorname{Re} \left(\dot{a}_k^* y'(kT + \tau) \right). \quad (7.16)$$

Структурная схема получившейся системы временной синхронизации показана на рис. 7.10.

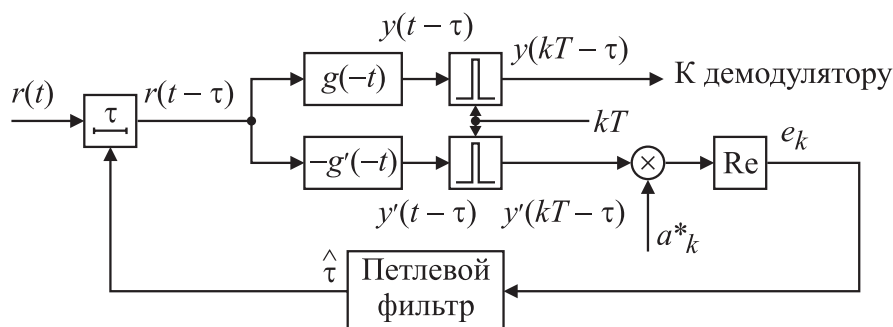


Рис. 7.10. Замкнутая система временной синхронизации с использованием данных

Такой вариант формирования сигнала ошибки требует наличия в системе дополнительного дифференцирующего фильтра, что увеличивает вычислительные затраты. Поэтому получили практическое распространение варианты алгоритма, использующие приближенную реализацию дифференцирования. Простейший способ — заменить производную $y'(t)$ конечной разностью: $y'(t) \approx (y(t + \delta) - y(t - \delta)) / (2\delta)$. Популярным является вариант этой формулы с $\delta = T/2$, что после удаления незначущего постоянного множителя дает следующий алгоритм формирования сигнала ошибки:

$$e_k = \operatorname{Re} \left(\dot{a}_k^* \left(y \left(kT + \frac{T}{2} + \tau_k \right) - y \left(kT - \frac{T}{2} + \tau_{k-1} \right) \right) \right). \quad (7.17)$$

Разные индексы у временного сдвига τ в (7.17) отражают тот факт, что отсчеты сигнала на выходе СФ формируются в замкнутой системе с обратной связью, а оценки τ обновляются лишь в моменты времени kT . Поэтому в тот момент, когда формируется отсчет $y(kT - T/2 + \tau_{k-1})$, последней доступной оценкой τ является значение, полученное в момент $(k-1)T$, т. е. τ_{k-1} . В момент же формирования отсчета $y(kT + T/2 + \tau_k)$ уже сформирована следующая оценка τ , т. е. τ_k .

7.5.3. Разомкнутый алгоритм без использования данных

В основе слепой временной синхронизации сигналов с линейной модуляцией лежит периодичность зависимости средней мощности выходного сигнала согласованного фильтра $y(t)$ от времени.

Исходя из (2.16), средняя мощность этого сигнала

$$\begin{aligned} P_{\text{cp}}(t) &= \overline{|y(t)|^2} = \overline{\left| \sum_{k=-\infty}^{\infty} a_k B_g(t - kT) \right|^2} = \\ &= \overline{\left(\sum_{k_1=-\infty}^{\infty} a_{k_1} B_g(t - k_1 T) \right) \left(\sum_{k_2=-\infty}^{\infty} a_{k_2} B_g(t - k_2 T) \right)^*} = \\ &= \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} \overline{a_{k_1} a_{k_2}^*} B_g(t - k_1 T) B_g^*(t - k_2 T). \end{aligned}$$

Так как передаваемые символы считаются статистически независимыми и имеющими нулевое среднее, в этой формуле остаются только слагаемые с $k_1 = k_2$:

$$P_{\text{cp}}(t) = \sigma_a^2 \sum_{k=-\infty}^{\infty} |B_g(t - kT)|^2. \quad (7.18)$$

Из получившегося выражения видно, что средняя мощность является периодической функцией времени с периодом T . Количество ненулевых слагаемых в (7.18) конечно и определяется удвоенной длительностью сигнальной посылки $g(t)$. Максимумы этой функции наблюдаются в моменты времени $t = kT$.

На рис. 7.11 в качестве примера показан вид функции $P_{\text{cp}}(t)$ в случае прямоугольной сигнальной посылки (сплошная линия) и при использовании

формирующего фильтра с косинусоидальным сглаживанием АЧХ и коэффициентом сглаживания 0,5 (штриховая линия).

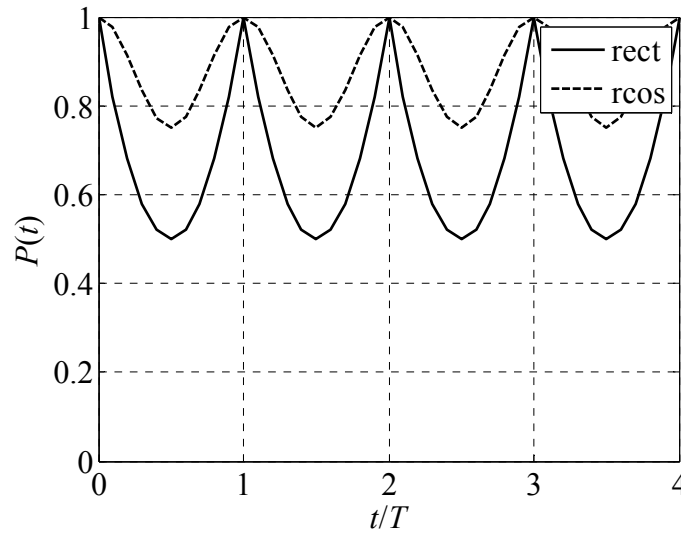


Рис. 7.11. Зависимость средней мощности сигнала от времени

Таким образом, для выполнения слепой временной синхронизации необходимо найти такой временной сдвиг τ , при котором средняя мощность отсчетов сигнала после согласованного фильтра будет максимальной:

$$\sum_{k=-\infty}^{\infty} |y(kT + \tau)|^2 \rightarrow \max_{\tau}. \quad (7.19)$$

Считая форму функции $P_{cp}(t)$ приближенно гармонической (это оправдано при использовании сглаженных сигнальных посылок, см. рис. 7.11), можно вычислить положение искомого максимума следующим образом:

$$\hat{\tau} = \frac{T}{2\pi} \arg \left(\int_{-\infty}^{\infty} |y(t)|^2 \exp \left(j2\pi \frac{t}{T} \right) dt \right). \quad (7.20)$$

Выражение (7.20) описывает распространенный метод разомкнутой слепой временной синхронизации.

7.5.4. Замкнутый алгоритм без использования данных

Чтобы получить сигнал ошибки для замкнутой системы синхронизации, как обычно, продифференцируем максимизируемое выражение (7.19):

$$\begin{aligned}
\varepsilon &= \frac{d}{d\tau} \left(\sum_{k=-\infty}^{\infty} |y(kT + \tau)|^2 \right) = \sum_{k=-\infty}^{\infty} \frac{d}{d\tau} (y(kT + \tau) y^*(kT + \tau)) = \\
&= \sum_{k=-\infty}^{\infty} y(kT + \tau) y'^*(kT + \tau) + y'(kT + \tau) y^*(kT + \tau) = \\
&= \sum_{k=-\infty}^{\infty} 2 \operatorname{Re} (y'(kT + \tau) y^*(kT + \tau)).
\end{aligned}$$

Отбрасывая постоянный множитель и возлагая усреднение по k на петлевой фильтр, получаем выражение для сигнала ошибки:

$$e_k = \operatorname{Re} (y^*(kT + \tau) y'(kT + \tau)). \quad (7.21)$$

Полученная формула отличается от (7.16) лишь тем, что вместо известных данных a_k в ней фигурирует просто значение сигнала на выходе СФ.

Структурная схема полученной системы синхронизации показана на рис. 7.12. Ее недостаток тот же, что и для схемы с использованием данных — необходимость в наличии дополнительного дифференцирующего фильтра. Можно упростить схему, заменив дифференцирование, как и ранее, конечной разностью. Полученная формула для сигнала ошибки в случае $\delta = T/2$ (см. формулу (7.17)) будет иметь вид

$$e_k = \operatorname{Re} \left(y^*(kT + \tau_{k-1}) \left(y \left(kT + \frac{T}{2} + \tau_k \right) - y \left(kT - \frac{T}{2} + \tau_{k-1} \right) \right) \right).$$

Она также отличается от (7.17) лишь тем, что вместо известных данных a_k в ней фигурирует просто значение сигнала на выходе СФ.

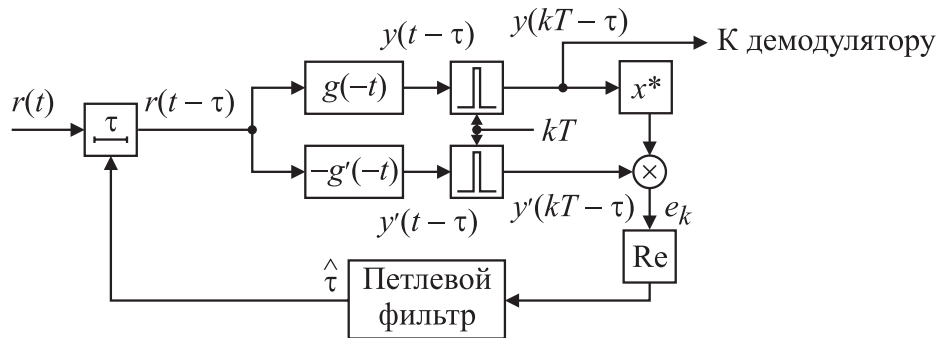


Рис. 7.12. Замкнутая схема слепой временной синхронизации

Следует отметить, что данный метод получен *эвристическим* путем, не опираясь на правило МП. Однако попытка более точной реализации МП-подхода приводит к аналогичному по своей сущности алгоритму, лишь использующему другой, более сложно рассчитываемый фильтр вместо дифференцирующего фильтра, необходимого для реализации выражения (7.21).

8. ПОМЕХОУСТОЙЧИВОЕ КОДИРОВАНИЕ

8.1. Основные принципы и термины

8.1.1. Идея помехоустойчивого кодирования

При передаче цифровой информации по каналу связи из-за наличия шумов и помех существует некоторая вероятность искажения передаваемых символов. Для обнаружения и/или исправления некоторых из возникающих ошибок может использоваться *помехоустойчивое кодирование* (Error-Correcting Coding, ECC; используется также термин «кодирование канала» — Channel Coding). Его идея заключается в том, что в передаваемое сообщение вносится контролируемая *избыточность*. Добавляемые избыточные символы связаны определенными математическими соотношениями с символами сообщения. Цель внесения избыточности — сделать возможные кодированные сообщения как можно больше отличающимися друг от друга. В результате внесения избыточности оказывается, что являются допустимыми не все возможные последовательности символов, а только те, что удовлетворяют использованному алгоритму кодирования. Именно этот факт дает возможность при приеме искаженного сообщения обнаружить ошибку и, возможно, исправить ее.

В данном разделе рассматривается прежде всего влияние помехоустойчивого кодирования на помехоустойчивость систем цифровой связи. Вопросам построения кодов, реализации процедур кодирования и декодирования посвящены многочисленные книги, такие, как, например, [8], [9].

8.1.2. Способы коррекции ошибок

В зависимости от того, используется в системе *обнаружение* или *исправление* ошибок с помощью помехоустойчивого кода, различают следующие варианты:

- *запрос повторной передачи* (Automatic Repeat reQuest, ARQ): с помощью помехоустойчивого кода выполняется только *обнаружение* ошибок, при их наличии производится запрос на повторную передачу пакета данных;
- *прямое исправление ошибок* (Forward Error Correction, FEC): производится *декодирование* помехоустойчивого кода, т. е. *исправление ошибок* с его помощью.

Возможен также гибридный вариант, когда делается попытка исправления ошибок и, если их число превышает некоторый порог, результат считается недостаточно достоверным и посылается запрос на повторную передачу.

8.1.3. Форма кривых помехоустойчивости для кодированных систем

Пусть в системе связи используется исправление ошибок с помощью помехоустойчивого кода. Не следует думать, что кривая помехоустойчивости

для кодированной системы при всех отношениях сигнал/шум будет идти ниже аналогичной кривой для системы без кодирования, использующей тот же вид модуляции. Для иллюстрации на рис. 8.1 показаны кривая помехоустойчивости для некодированной системы с модуляцией ФМ-2 и кривая для системы с тем же видом модуляции, но использующей код Хэмминга (63, 57) с декодированием жестких решений (смысл этих терминов будет ясен далее).

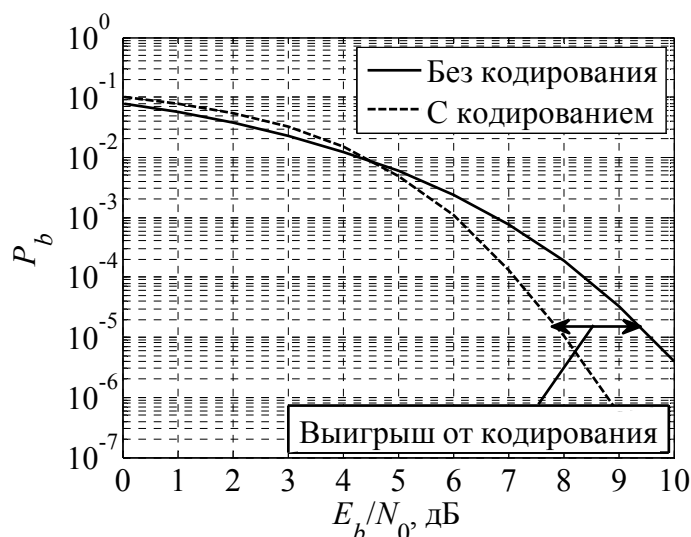


Рис. 8.1. Взаимное расположение кривых помехоустойчивости для кодированной и некодированной систем связи, использующих один и тот же вид модуляции

Из графиков видно, что в определенном диапазоне отношений сигнал/шум помехоустойчивость кодированной системы хуже, чем некодированной. Дело в том, что при введении в систему помехоустойчивого кода начинают работать 2 механизма. С одной стороны, из-за введения в сообщение избыточности по каналу связи передается больше бит и, следовательно, уменьшается энергия, приходящаяся на один канальный бит. Из-за этого вероятность ошибки в демодулированном, но еще не декодированном сообщении оказывается выше, чем при отсутствии кодирования (под энергией бита в отношении E_b/N_0 для кодированных систем имеется в виду средняя энергия, приходящаяся на один *информационный* бит, т. е. бит исходного (некодированного) сообщения). С другой стороны, при декодировании помехоустойчивого кода исправляется часть произошедших в канале ошибок, так что вероятность ошибки уменьшается. При разных отношениях сигнал/шум соотношение между этими двумя эффектами оказывается различным. При больших отношениях сигнал/шум снижение вероятности ошибки за счет свойств кода оказывается более существенным, чем рост вероятности ошибки из-за уменьшения энергии канальных бит, при малых отношениях сигнал/шум все происходит наоборот. В результате кривые помехоустойчивости для кодированной и некодированной систем пересекаются. Для используемых на прак-

тике кодов точка пересечения, как правило, расположена в области довольно высоких вероятностей ошибки (единицы или даже десятки процентов).

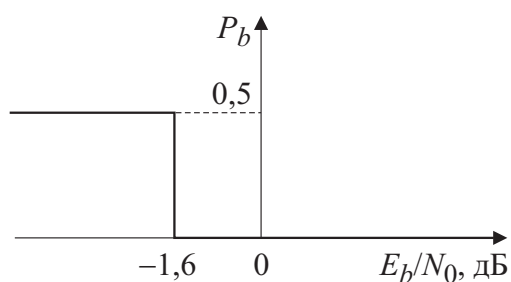


Рис. 8.2. Предельная форма кривой помехоустойчивости при неограниченном увеличении корректирующей способности кода

Итак, при достаточно больших отношениях сигнал/шум на бит использование помехоустойчивого кода позволяет получить ту же вероятность ошибки при меньшем отношении сигнал/шум на бит, чем в некодированной системе. Этот энергетический выигрыш называется *выигрышем от кодирования* (coding gain, см. рис. 8.1).

При неограниченном росте корректирующей способности кода кривая помехоустойчивости будет стремиться к своей предельной форме, которая следует из формулы Шеннона (1.4): если отношение сигнал/шум на бит больше предела Шеннона ($-1,6$ дБ, см. (1.5)), вероятность ошибки стремится к нулю, в противном случае она стремится к $1/2$. Получающаяся в результате предельная форма кривой помехоустойчивости показана на рис. 8.2.

8.1.4. Классификация кодов

Помехоустойчивые коды можно классифицировать по разным признакам. Приведем некоторые из них.

Блочные и сверточные коды. Наиболее принципиальным является деление помехоустойчивых кодов на блочные и сверточные. При использовании *блокового кода* (block code) исходное сообщение делится на фрагменты фиксированной длины — *блоки* и каждый блок кодируется независимо, т. е. получающееся в результате кодирования *кодированное слово* (codeword) не зависит от содержания предыдущих блоков. В противоположность этому кодер для *сверточного кода* (convolutional code) представляет собой устройство с *памятью*. Символ, поступающий на вход кодера, по определенному закону изменяет внутреннее состояние кодера. Генерируемое кодером кодированное сообщение (оно называется *кодированной последовательностью* — code sequence) зависит как от текущего входного символа, так и от внутреннего состояния кодера. Таким образом, сообщение не делится на блоки и результат кодирования зависит не только от текущего фрагмента сообщения, но и от внутреннего состояния кодера, т. е., в конечном счете, от всей предыстории его работы.

Систематические и несистематические коды. В *систематическом* коде результат кодирования (кодированное слово или кодированная последовательность) содержит исходное сообщение в неизменном виде, так что символы на вы-

ходе кодера можно разделить на *информационные* и *избыточные* (добавленные в результате кодирования). На выходе кодера, использующего *несистематический* код, символы исходного сообщения в явном виде не присутствуют. С точки зрения теории кодирования не имеет большого значения, является код систематическим или нет. Однако систематический характер кода упрощает некоторые аспекты практической реализации кодированных систем. Например, если ошибок в кодовом слове не обнаружено, для извлечения из него информационной части в случае систематического кода достаточно отбросить избыточные символы.

Двоичные и недвоичные коды. Согласно алфавиту, которому принадлежат символы сообщения и результата его кодирования, различают *двоичные* и *недвоичные* коды. Большинство используемых на практике кодов является двоичными; распространенным семейством недвоичных кодов являются блочные коды Рида—Соломона.

Линейные и нелинейные коды. *Линейными* называются коды, в которых произвольная линейная комбинация кодовых слов дает также допустимое кодовое слово. В противном случае код называется *нелинейным*. В случае систематического линейного кода можно сказать, что избыточные символы связаны с информационными символами линейными соотношениями.

8.1.5. Жесткие и мягкие решения

Термины «жесткие» и «мягкие» решения описывают способ взаимодействия демодулятора и декодера.

Жесткими решениями (hard decisions) называется режим работы системы «демодулятор—декодер», при котором демодулятор принимает окончательные решения о значениях (0 или 1) бит, принятых из канала связи, и только эти решения передает декодеру помехоустойчивого кода.

При использовании жестких решений теряется информация о том, насколько *достоверным* является решение, принятое демодулятором о значении того или иного бита. Так, при использовании линейной модуляции эта достоверность определяется удаленностью принятого отсчета от границы областей принятия решения в пользу разных точек сигнального созвездия. Декодирование с *мягкими решениями* (soft decisions) предполагает использование этой информации.

На практике декодирование мягких решений чаще применяется для сверточных кодов, подробнее оно будет рассмотрено в 8.3.5.

8.2. Блочные коды

8.2.1. Основные параметры

Как уже отмечалось, при использовании блочных кодов исходное сообщение делится на блоки фиксированного размера. Число символов в блоке

называется *длиной сообщения* (message length) и обычно обозначается буквой k . В результате кодирования блока получается набор из $n > k$ символов, называемый *кодовым словом* (codeword). Число добавляемых избыточных символов, таким образом, равно $n - k$.

Эти «размерные» характеристики блочного кода обычно записывают в круглых скобках через запятую: (n, k) . Таким образом, например, термин «код $(7, 4)$ » означает, что речь идет о коде, при использовании которого исходное сообщение делится на блоки длиной 4 символа, а размер получаемых кодовых слов составляет 7 символов. К каждому блоку добавляется $7 - 4 = 3$ избыточных символа.

Из-за внесения избыточности возрастает объем данных, передаваемых по каналу связи. Доля «полезных» данных составляет при этом $r = k/n$. Данное отношение называется *скоростью кодирования* (code rate).

Важнейшей характеристикой блочного кода является *кодовое расстояние* (code distance) — мера «непохожести» возможных кодовых слов. Для измерения расстояния между кодовыми словами могут использоваться различные *метрики*. Наибольшее распространение в теории помехоустойчивого кодирования получила метрика Хэмминга, при использовании которой расстояние между двумя кодовыми словами равно числу несовпадающих символов в них. Кодовым расстоянием называется минимальное из попарных расстояний между всеми возможными кодовыми словами.

Кодовое расстояние обозначается как d_{\min} и иногда указывается в качестве третьей характеристики кода в круглых скобках: (n, k, d_{\min}) .

От кодового расстояния зависит *корректирующая способность* кода, т. е. возможность обнаруживать и/или исправлять ошибки с его помощью.

Очевидно, что если в одном блоке искажается не более чем $d_{\min} - 1$ символов, то кодовое слово превращается в недопустимую («неправильную») комбинацию символов, и такая ошибка будет обнаружена. Если же окажутся искаженными d_{\min} или более символов, в результате *может* получиться другое допустимое кодовое слово и ошибка останется незамеченной. Таким образом, максимальное число искаженных символов в блоке (его называют *кратностью ошибки*), при котором ошибка *гарантированно* обнаруживается, составляет $v = d_{\min} - 1$.

Для определения возможностей кода по *исправлению* ошибок необходимо решить, что следует делать в том случае, когда принятый блок не является допустимым кодовым словом. Логичным решением будет выбор того из допустимых кодовых слов, которое максимально близко (в смысле расстояния Хэмминга) к принятой комбинации символов. Данный метод является декодированием по правилу МП.

Если в блоке оказались искаженными менее чем $d_{\min}/2$ символов, то исходное кодовое слово будет ближайшим к принятой комбинации символов и ошибка будет исправлена. Если же число искаженных символов в блоке превышает $d_{\min}/2$, принятая комбинация символов окажется ближе уже к другому кодовому слову и результат декодирования окажется неправильным.

При четном d_{\min} и кратности ошибки, равной $d_{\min}/2$, возможно, что найдутся 2 кодовых слова, находящиеся на равных расстояниях от принятой комбинации символов, поэтому результат декодирования может быть как правильным, так и ошибочным.

Таким образом, кратность ошибки, *гарантированно* исправляемой при декодировании, составляет $t = [(d_{\min} - 1)/2]$. Квадратными скобками здесь обозначено выделение целой части числа.

При использовании кода возможна комбинация обнаружения и исправления ошибок. При этом можно исправлять $t' \leq t$ ошибок и *дополнительно к ним* обнаруживать еще $v' \leq v$ ошибок. Реализуется такой режим следующим образом: если принятая комбинация символов находится на хэмминговом расстоянии не более чем t' от одного из кодовых слов, ошибка исправляется, в противном случае выдается сигнал об обнаружении ошибки кратности больше t' . Возможные сочетания значений t' и v' должны удовлетворять следующему неравенству: $2t' + v' \leq d_{\min} - 1$. Смысл этого неравенства иллюстрирует рис. 8.3.

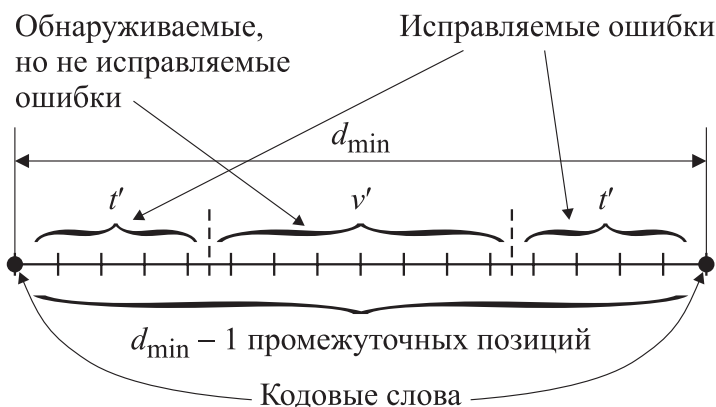


Рис. 8.3. Связь между числом исправляемых и обнаруживаемых ошибок для блочного кода

Один и тот же код можно использовать для обнаружения ошибок, исправления ошибок или для комбинации того и другого. Так как обнаружить можно *больше* ошибок, чем исправить, то вероятность ошибки в системе с запросом повтора передачи будет меньше, чем в системе с исправлением ошибок, использующей тот же код.

8.2.2. Общие принципы кодирования и декодирования

Линейный блочный код в общем случае описывается *порождающей матрицей* (generator matrix). При таком описании кодирование блока (вектора) производится умножением его на порождающую матрицу. В качестве примера рассмотрим систематический двоичный код Хэмминга (7, 4), формирование избыточных символов для которого описывается следующими формулами:

$$\begin{aligned} y_5 &= x_1 \oplus x_2 \oplus x_3, \\ y_6 &= x_2 \oplus x_3 \oplus x_4, \\ y_7 &= x_1 \oplus x_2 \oplus x_4. \end{aligned} \quad (8.1)$$

Здесь $x_1 \dots x_4$ — биты блока сообщения; $y_5 \dots y_7$ — избыточные биты кодового слова. Поскольку код является систематическим, первые 4 бита кодового слова повторяют блок исходного сообщения: $y_k = x_k$ для $k = 1, \dots, 4$.

Все 16 кодовых слов данного кода приведены в табл. 8.1.

Табл. 8.1. Код Хэмминга (7, 4):
кодовые слова

x_1	x_2	x_3	x_4			
y_1	y_2	y_3	y_4	y_5	y_6	y_7
0	0	0	0	0	0	0
1	0	0	0	1	0	1
0	1	0	0	1	1	1
1	1	0	0	0	1	0
0	0	1	0	1	1	0
1	0	1	0	0	1	1
0	1	1	0	0	0	1
1	1	1	0	1	0	0
0	0	0	1	0	1	1
1	0	0	1	1	1	0
0	1	0	1	1	0	0
1	1	0	1	0	0	1
0	0	1	1	1	0	1
1	0	1	1	0	0	0
0	1	1	1	0	1	0
1	1	1	1	1	1	1

Табл. 8.2. Код Хэмминга (7, 4): таблица
декодирования

Синдром			Вектор ошибки						
s_1	s_2	s_3	e_1	e_2	e_3	e_4	e_5	e_6	e_7
1	0	1	1	0	0	0	0	0	0
1	1	1	0	1	0	0	0	0	0
1	1	0	0	0	1	0	0	0	0
0	1	1	0	0	0	1	0	0	0
1	0	0	0	0	0	0	1	0	0
0	1	0	0	0	0	0	0	1	0
0	0	1	0	0	0	0	0	0	1

В матричном виде процесс кодирования можно представить следующим образом:

$$\mathbf{y} = \mathbf{x} \mathbf{G}, \quad (8.2)$$

где $\mathbf{x} = [x_1 \ x_2 \ x_3 \ x_4]$; $\mathbf{y} = [y_1 \ y_2 \ y_3 \ y_4 \ y_5 \ y_6 \ y_7]$, а \mathbf{G} — порождающая матрица размером $k \times n$, имеющая в данном случае вид

$$\mathbf{G} = \left[\begin{array}{cccc|ccc} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{array} \right].$$

Математические операции при использовании двоичных кодов производятся *по модулю 2*.

Помимо порождающей, существует *проверочная матрица* кода \mathbf{H} . Эта матрица реализует проверку соотношений, использованных при формировании кодового слова. В данном примере из (8.1) с учетом систематического характера кода получаются следующие проверочные соотношения:

$$y_1 \oplus y_2 \oplus y_3 \oplus y_5 = 0,$$

$$y_2 \oplus y_3 \oplus y_4 \oplus y_6 = 0,$$

$$y_1 \oplus y_2 \oplus y_4 \oplus y_7 = 0.$$

В матричном виде это можно записать следующим образом:

$$\mathbf{y} \mathbf{H}^T = \mathbf{0}, \quad (8.3)$$

где \mathbf{H} — проверочная матрица размером $(n - k) \times n$. Для рассматриваемого кода она выглядит следующим образом:

$$\mathbf{H} = \left[\begin{array}{cccc|ccc} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{array} \right].$$

Соотношение (8.3) используется для обнаружения ошибок в кодовом слове.

Подставив (8.2) в (8.3), получаем соотношение $\mathbf{x} \mathbf{G} \mathbf{H}^T = \mathbf{0}$, которое должно выполняться при любом информационном блоке \mathbf{x} . Отсюда следует, что произведение порождающей и проверочной матриц дает нулевую матрицу: $\mathbf{G} \mathbf{H}^T = \mathbf{0}$.

Если умножение кодированного блока на проверочную матрицу не дает нулевого вектора, то полученный результат \mathbf{s} (его называют *синдромом* — syndrome) позволяет определить, какие именно символы, скорее всего, были искажены в процессе передачи и чему равны их первоначальные значения. Если искаженное кодовое слово представить в виде $\mathbf{y}' = \mathbf{y} + \mathbf{e}$, где \mathbf{e} — вектор ошибок, то из (8.3) следует, что

$$\mathbf{y}' \mathbf{H}^T = \mathbf{e} \mathbf{H}^T = \mathbf{s}. \quad (8.4)$$

Декодирование линейного блочного кода, таким образом, можно осуществить с помощью таблицы, в которой для каждого значения синдрома указан соответствующий вектор ошибок.

Система уравнений (8.4) является недоопределенной, в качестве ее решений выбираются значения вектора \mathbf{e} с минимальным числом ненулевых

элементов (так как в АБГШ-канале чем больше кратность ошибки, тем меньше вероятность ее возникновения).

Табл. 8.2 иллюстрирует декодирование кода Хэмминга (7, 4), позволяющего исправить одну ошибку.

Табличное декодирование можно реализовать на практике только при сравнительно небольших значениях n и k . Поэтому основным направлением развития теории помехоустойчивого кодирования является поиск семейств кодов, декодирование которых может быть реализовано алгебраическим путем, т. е. с помощью выполнения некоторых математических операций [8], [9].

8.2.3. Весовой спектр и вероятность необнаружения ошибки

Важной характеристикой двоичного линейного блочного кода является его *весовой спектр*. *Весом* кодового слова называется количество единичных бит в нем, а *весовой спектр* определяется как совокупность целых чисел $\{A_d\}$, каждое из которых показывает количество кодовых слов с весом d . Очевидно, что $A_d = 0$ для $d = 1, \dots, d_{\min} - 1$. Кроме того, линейный код обязательно содержит кодовое слово, состоящее из одних нулей, поэтому $A_0 = 1$.

Табл. 8.3. Код Хэмминга (7, 4):
весовой спектр

d	0	1	2	3	4	5	6	7
A_d	1	0	0	7	7	0	0	1

Анализ табл. 8.1 позволяет получить весовой спектр для кода Хэмминга (7, 4). Этот спектр приведен в табл. 8.3.

Весовой спектр используется, в частности, для расчета вероятности необнаружения ошибки с помощью двоичного линейного блочного кода. Ошибка не обнаруживается, если в результате нее одно кодовое слово превращается в другое. Если расстояние Хэмминга между правильным и неправильным кодовыми словами равно d , это означает, что произошло d битовых ошибок в конкретных позициях, а оставшиеся $n - d$ бит остались неизменными. Вероятность этого события равна

$$p^d (1 - p)^{n-d}, \quad (8.5)$$

где p — вероятность ошибочного приема бита кодового слова. Она рассчитывается исходя из используемого вида модуляции с учетом того, что энергия бит в канале связи уменьшается (по сравнению с некодированной системой) с коэффициентом, равным скорости кодирования r .

События перепутывания правильного кодового слова с разными другими кодовыми словами являются несовместными, поэтому их вероятности можно суммировать. Кроме того, так как рассматриваемый код является линейным, усреднять результаты расчета по разным кодовым словам нет необходимости, поскольку эти результаты окажутся одинаковыми. Поэтому в качестве «правильного» при расчете можно использовать любое кодовое слово, удобнее всего выбрать слово, состоящее из одних нулей. С учетом изложен-

ного получаем следующую формулу для вероятности необнаружения возникших ошибок:

$$P_{\text{необн}} = \sum_{d=d_{\min}}^n A_d p^d (1-p)^{n-d}. \quad (8.6)$$

Для кода Хэмминга (7, 4) с помощью (8.6) и данных из табл. 8.3 получаем следующую формулу для вероятности необнаружения ошибки:

$$P_{\text{необн}} = 7p^3(1-p)^{7-3} + 7p^4(1-p)^{7-4} + p^7(1-p)^{7-7} = 7p^3(1-p)^3 + p^7.$$

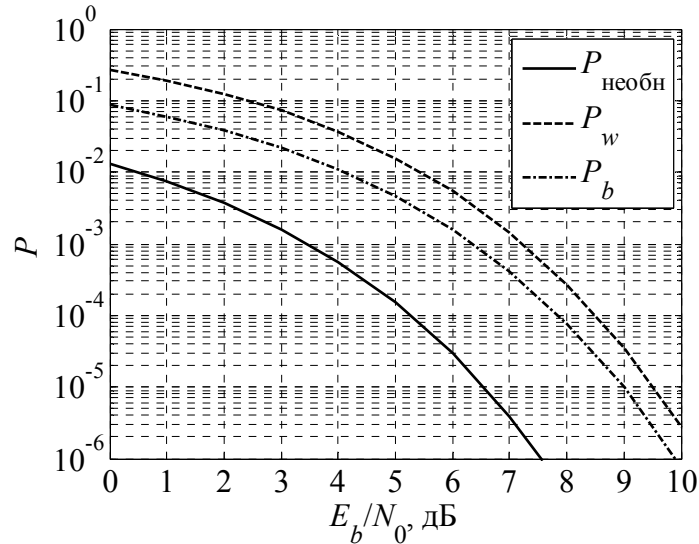


Рис. 8.4. Вероятности необнаружения ошибки, неправильного декодирования кодового слова и битовой ошибки для кода Хэмминга (7, 4)

График получившейся зависимости показан на рис. 8.4 (сплошная линия). Подразумевалось, что используется модуляция ФМ-2, так что вероятность p ошибочного приема бита кодового слова

$$p = Q\left(\sqrt{\frac{4}{7} \frac{2E_b}{N_0}}\right). \quad (8.7)$$

Множитель $4/7$, равный скорости кодирования, учитывает уменьшение энергии канальных бит из-за внесения избыточности в сообщение.

8.2.4. Вероятность ошибки декодирования

Оценить вероятность P_w ошибки декодирования кодового слова можно следующим образом. Если код гарантированно исправляет t ошибок, то ошибка декодирования может возникнуть, если в кодовом слове произошло больше чем t ошибок. Вероятность того, что произойдет d битовых ошибок в конкретных позициях кодового слова, дается приведенной ранее формулой (8.5). Количество различных вариантов ошибки кратности d равно числу со-

четаний из n по d , C_n^d . Таким образом, верхняя граница для вероятности ошибки может быть записана как

$$P_w \leq \sum_{d=t+1}^n C_n^d p^d (1-p)^{n-d} = 1 - \sum_{d=0}^t C_n^d p^d (1-p)^{n-d}. \quad (8.8)$$

Формула (8.8) — это именно верхняя граница, так как код в общем случае может исправлять *некоторые* ошибки кратности больше t . Если код исправляет *все* ошибки кратности до t включительно и *ни одной* ошибки большей кратности, такой код называется *совершенным* (perfect). Совершенные коды имеют максимально возможную корректирующую способность t при данных n и k ; для них неравенство (8.8) превращается в равенство.

Код Хэмминга (7, 4) имеет кодовое расстояние $d_{\min} = 3$ и, таким образом, позволяет исправлять одну ошибку в кодовом слове, так что $t = 1$. Кроме того, данный код является совершенным, поэтому (8.8) превращается в равенство. С учетом изложенного для данного кода получаем

$$P_w = 1 - \sum_{d=0}^1 C_7^d p^d (1-p)^{7-d} = 21p^2 - 70p^3 + 105p^4 - 84p^5 + 35p^6 - 6p^7.$$

График получившейся зависимости показан на рис. 8.4 штриховой линией. Как и для расчета вероятности необнаружения ошибки, предполагалось использование модуляции ФМ-2 (см. (8.7)). Графики количественно демонстрируют упоминавшийся ранее факт, что при использовании одного и того же кода вероятность пропуска (необнаружения) ошибки существенно меньше, чем вероятность ошибки декодирования.

Что касается вероятности *битовой* ошибки, она зависит от того, как именно устроен код (подобно тому как вероятность битовой ошибки при многопозиционной модуляции зависит от того, как именно распределены битовые комбинации по точкам созвездия). Приближенное соотношение, приводимое в [1, 6.5.3], имеет вид

$$\begin{aligned} P_b &\approx \frac{1}{n} \sum_{d=t+1}^n d C_n^d p^d (1-p)^{n-d} = \sum_{d=t+1}^n C_{n-1}^{d-1} p^d (1-p)^{n-d} = \\ &= p \left(1 - \sum_{d=0}^{t-1} C_{n-1}^d p^d (1-p)^{n-1-d} \right). \end{aligned} \quad (8.9)$$

Эта формула получается исходя из следующих предположений:

- если число произошедших ошибок превосходит t , процесс декодирования *в среднем не изменяет* число битовых ошибок в кодовом слове;
- эти ошибки равномерно распределены между систематической и избыточной частями кодового слова.

Для кода Хэмминга (7, 4) по формуле (8.9) получаем

$$P_b \approx p \left(1 - \sum_{d=0}^{1-1} C_6^d p^d (1-p)^{6-d} \right) = p \left(1 - (1-p)^6 \right).$$

Соответствующий график показан на рис. 8.4 штрихпунктирной линией. Из графика видно, что вероятность битовой ошибки декодирования меньше, чем вероятность неправильного декодирования кодового слова.

8.3. Сверточные коды

8.3.1. Основные параметры

Как уже отмечалось, кодер для сверточного кода представляет собой устройство с памятью. Как и для блочного кода, символы подаются на вход кодера группами по k символов, и в результате кодирования одной такой группы на выходе кодера появляется $n > k$ выходных символов (символов кодированного сообщения). Скорость кодирования, также аналогично блочным кодам, составляет $r = k/n$. Однако, в отличие от блочных кодов, символы выходной группы зависят не только от символов входной группы, но и от *внутреннего состояния* кодера, которое меняется по определенному закону при поступлении на вход очередной группы символов. Можно, таким образом, считать, что блочный код — частный случай сверточного кода, получаемый при полном отсутствии у кодера внутренней памяти.

В отличие от блочных кодов, для сверточных кодов числа k и n , как правило, невелики. Так, для многих используемых на практике сверточных кодов $k = 1$, так что скорость кодирования $r = 1/n$.

Исходя из изложенного, работу сверточного кодера можно описать с помощью двух функций f и g следующим образом:

$$\begin{aligned} \mathbf{s}_{m+1} &= f(\mathbf{s}_m, \mathbf{x}_m), \\ \mathbf{y}_m &= g(\mathbf{s}_m, \mathbf{x}_m), \end{aligned} \tag{8.10}$$

где m — номер группы символов; \mathbf{s} — вектор внутреннего состояния кодера; \mathbf{x} и \mathbf{y} — группы символов на входе и выходе кодера соответственно.

Функция f описывает эволюцию внутреннего состояния \mathbf{s} под действием кодируемых данных \mathbf{x} , функция g отвечает за формирование кодированной последовательности. Эти функции могут быть как линейными, так и нелинейными. Строго говоря, термин «сверточный код» справедлив только в линейном случае, так как нелинейный код не описывается математической операцией свертки. Однако в литературных источниках последнего времени прослеживается тенденция использовать данный термин и для нелинейных кодов с памятью.

Количество групп входных символов (включая текущую группу), от которых зависит результат кодирования, называется *кодовым ограничением* (code constraint).

Минимальное расстояние Хэмминга между кодированными последовательностями называется *свободным расстоянием* (free distance) сверточного кода, далее оно будет обозначаться как d_{free} .

8.3.2. Кодирование и декодирование

Кодер для сверточного кода представляет собой один или несколько регистров сдвига, в которые поступают входные символы. Выходные символы формируются в результате выполнения математических операций над содержимым регистров.

В *рекурсивных* сверточных кодах регистры охвачены обратными связями.

В качестве примера на рис. 8.5 показана схема кодера для кода с кодовым ограничением 3 и скоростью кодирования $r = 1/2$.

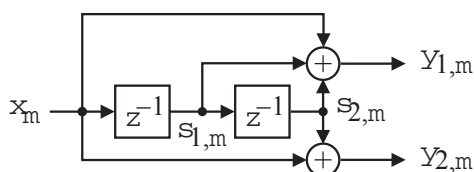


Рис. 8.5. Пример кодера для сверточного кода

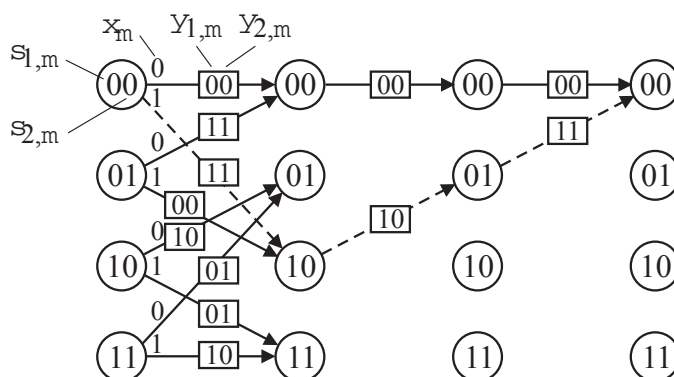


Рис. 8.6. Решетчатая диаграмма и свободное расстояние для кода, соответствующего рис. 8.5

Для этого кода $k = 1$, $n = 2$. Вектор внутреннего состояния состоит из двух элементов $s_{1,m}$ и $s_{2,m}$ — это значения, записанные в две ячейки памяти.

Функции (8.10) для данного примера имеют следующий вид:

$$\begin{aligned} s_{1,m+1} &= x_m, \\ s_{2,m+1} &= s_{1,m}, \\ y_{1,m} &= x_m \oplus s_{1,m} \oplus s_{2,m}, \\ y_{2,m} &= x_m \oplus s_{2,m}. \end{aligned}$$

Помимо схемы и функций (8.10) сверточный код можно описать различными способами: диаграммой состояний, решетчатой диаграммой, таблицей переходов, порождающими полиномами (в линейном случае). Решетчатая диаграмма для рассматриваемого примера кода показана на рис. 8.6.

Помимо собственно решетки переходов на рис. 8.6 показана пара путей, находящихся на минимальном хэмминговом расстоянии друг от друга. Один путь соответствует нулевой последовательности ($x_m = s_{1,m} = s_{2,m} = y_{1,m} = y_{2,m}$ для всех m , такая последовательность является допустимой для любого ли-

нейного кода), второй отходит от нее на протяжении трех тактов (этот путь показан на рисунке штриховыми стрелками). Данному пути соответствует кодированная последовательность 11-10-11, содержащая 5 единичных бит. Таким образом, свободное расстояние данного кода $d_{\text{free}} = 5$.

Декодирование сверточного кода по методу МП производится с помощью алгоритма Витерби. Используя терминологию обобщенного описания этого алгоритма из 1.6, можно считать, что в данном случае «города» — это внутренние состояния кодера, «дороги» — это переходы из одного внутреннего состояния кодера в другое, возможные согласно логике работы кодера. «Расстояния» — это метрики, используемые для определения степени близости принятого сигнала к возможным (опять-таки, согласно логике работы кодера) последовательностям символов. Цель алгоритма — отыскать допустимую кодированную последовательность, которая расположена максимально близко к принятому сигналу. Используемая для расчета расстояний метрика зависит от типа решений на выходе демодулятора. При жестких решениях это метрика Хэмминга, при мягких — евклидово расстояние между соответствующими сигналами.

8.3.3. Оценка вероятности ошибки

На каждом такте своей работы декодер Витерби сравнивает метрики путей и отбирает выжившие (см. 1.6), поэтому вероятность ошибки определяется вероятностью перепутать правильный путь по решетчатой диаграмме с каким-нибудь другим.

Рассмотрим две допустимые кодированные последовательности, одна из которых правильная, другая — ошибочная. Если код линейный, в качестве правильной можно рассматривать произвольную кодированную последовательность, удобно считать, что она полностью состоит из нулевых бит. Ошибочных последовательностей может быть много, мысленно пронумеруем их и будем считать, что рассматриваемая ошибочная последовательность имеет номер m .

Пусть вероятность перепутать правильную последовательность с m -й неправильной равна P_m , а число неправильных информационных бит, которые попадут на выход декодера при выборе этой неправильной последовательности, равно b_m . Тогда можно записать аддитивную верхнюю границу для вероятности битовой ошибки, усреднив число возникающих битовых ошибок по всем ошибочным последовательностям:

$$P_b \leq \sum_m b_m P_m. \quad (8.11)$$

Суммирование в этой формуле осуществляется по всем ошибочным кодированным последовательностям.

Теперь заметим, что P_m зависит только от расстояния Хэмминга d между правильной и ошибочной последовательностями. Группируя в (8.11) слагаемые для одинаковых d , получаем

$$P_b \leq \sum_{d=d_{\text{free}}}^{\infty} B_d P(d). \quad (8.12)$$

Здесь $P(d)$ — вероятность перепутывания кодированных последовательностей, находящихся на хэмминговом расстоянии d друг от друга; B_d — суммарное число ошибочных информационных бит во всех кодированных последовательностях, находящихся на хэмминговом расстоянии d от правильной (состоящей из одних нулей) последовательности.

С ростом d вероятность $P(d)$ стремится к нулю, поэтому при использовании (8.12) можно ограничиться конечным числом слагаемых.

Набор коэффициентов $\{B_d\}$, $d \geq d_{\text{free}}$, называется *дистанционным спектром* сверточного кода. Он может быть найден просмотром решетки переходов кода с использованием переборного алгоритма. Кроме того, в литературе (см., например, [1, 7.4.1]) приводятся методы определения дистанционного спектра аналитическим путем, исходя из структуры кодера.

Табл. 8.4. Дистанционный спектр кода, соответствующего рис. 8.5 и 8.6

d	5	6	7	8	9	10	11	12	13	14
B_d	1	4	12	32	80	192	448	1024	2304	5120

Для рассматриваемого примера (см. рис. 8.5 и 8.6) дистанционный спектр для расстояний d от 5 до 14 приведен в табл. 8.4.

8.3.4. Вероятность перепутывания путей при жестких решениях

Рассчитаем, чему равна вероятность $P(d)$ при использовании жестких решений. Для перепутывания последовательностей, находящихся на хэмминговом расстоянии d друг от друга, необходимо, чтобы в тех d битах, которыми отличаются последовательности, произошло более чем $d/2$ битовых ошибок (если d четно, то при возникновении ровно $d/2$ ошибок возможен выбор как правильной, так и неправильной последовательностей с вероятностью 50 %). Вероятность такого события

$$P(d) = \begin{cases} \sum_{i=\frac{d+1}{2}}^d C_d^i p^i (1-p)^{d-i}, & d \text{ нечетно,} \\ \sum_{i=\frac{d}{2}+1}^d C_d^i p^i (1-p)^{d-i} + \frac{1}{2} C_d^{d/2} p^{d/2} (1-p)^{d/2}, & d \text{ четно,} \end{cases}$$

где p — вероятность ошибочного приема бита кодовой последовательности, рассчитываемая аналогично случаю блочных кодов (см. 8.2.3). Так, при использовании модуляции ФМ-2

$$p = Q\left(\sqrt{\frac{2E_b}{N_0}} r\right),$$

где r — скорость кодирования.

8.3.5. Вероятность перепутывания путей при мягких решениях

В общем случае использование мягких решений предполагает максимизацию функции правдоподобия, т. е. совместной условной плотности вероятности для наблюдаемых отсчетов сигнала \dot{r}_k в предположении, что передавались различные допустимые кодированные последовательности. В АБГШ-канале отсчеты шума независимы, так что условные ПВ $p(\dot{r}_k | \dot{a}_k)$ для разных отсчетов перемножаются, а их логарифмы суммируются. Таким образом, правило мягкого декодирования имеет вид

$$\sum_k \ln p(\dot{r}_k | \dot{a}_k) \rightarrow \max_{\{\dot{a}_k\}}. \quad (8.13)$$

Максимизация в этой формуле производится по всем допустимым кодированным последовательностям.

Неудобство формулы (8.13) заключается в том, что каждое слагаемое в ней зависит как от принятого сигнала $\{\dot{r}_k\}$, так и от предполагаемой кодированной последовательности $\{\dot{a}_k\}$. Удобно было бы каким-либо образом разделить используемые величины, чтобы метрики, связанные с сигналом, вычислялись независимо от предположений о последовательности $\{\dot{a}_k\}$, а эти предположения формировались декодером в виде не зависящих от принятого сигнала числовых коэффициентов.

Чтобы добиться этого, прежде всего предположим, что используется бинарная модуляция, так что $a_k = \pm 1$. Теперь умножим (8.13) на 2 и вычтем из каждого слагаемого сумму логарифмов условных ПВ для случаев, когда переданный символ a_k был равен +1 и -1. Это преобразование, очевидно, является монотонным, поэтому правомерно искать максимум и от преобразованной указанным образом функции:

$$\sum_k (2 \ln p(\dot{r}_k | a_k) - \ln p(\dot{r}_k | +1) - \ln p(\dot{r}_k | -1)) \rightarrow \max_{\{a_k\}}. \quad (8.14)$$

Теперь заметим, что суммируемые слагаемые можно записать следующим образом:

$$2 \ln p(\dot{r}_k | a_k) - \ln p(\dot{r}_k | +1) - \ln p(\dot{r}_k | -1) =$$

$$= \begin{cases} \ln p(\dot{r}_k | +1) - \ln p(\dot{r}_k | -1), & a_k = +1, \\ \ln p(\dot{r}_k | -1) - \ln p(\dot{r}_k | +1), & a_k = -1 \end{cases} = a_k \ln \frac{p(\dot{r}_k | +1)}{p(\dot{r}_k | -1)}.$$

С учетом этого правило (8.14) принимает вид

$$\sum_k a_k \ln \frac{p(\dot{r}_k | +1)}{p(\dot{r}_k | -1)} \rightarrow \max_{\{a_k\}}. \quad (8.15)$$

Поставленная цель достигнута — в левой части (8.15) кодированная последовательность представлена в виде числовых коэффициентов $a_k = \pm 1$, а сигнал — в виде так называемых *логарифмических отношений правдоподобия* (Log-Likelihood Ratio, LLR), зависящих только от его отсчетов и рассчитываемых как

$$\text{LLR}(\dot{r}_k) = \ln p(\dot{r}_k | +1) - \ln p(\dot{r}_k | -1) = \ln \frac{p(\dot{r}_k | +1)}{p(\dot{r}_k | -1)}. \quad (8.16)$$

Эти логарифмические отношения правдоподобия и представляют собой *мягкие решения* на выходе демодулятора.

Заметим, что жесткие решения представляют собой результат *сравнения* условных ПВ и, таким образом, могут быть рассчитаны как *знаки* мягких решений, $\text{sign}(\text{LLR}(\dot{r}_k))$.

С учетом введенного обозначения (8.16) правило декодирования с мягкими решениями (8.14) можно окончательно записать следующим образом:

$$\sum_k a_k \text{LLR}(\dot{r}_k) \rightarrow \max_{\{a_k\}}.$$

Рассчитаем, чему равны логарифмические отношения правдоподобия при использовании модуляции ФМ-2. Входящие в (8.16) условные ПВ в данном случае равны:

$$p(\dot{r}_k | +1) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|\dot{r}_k - 1|^2}{2\sigma^2}\right), \quad p(\dot{r}_k | -1) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|\dot{r}_k + 1|^2}{2\sigma^2}\right), \quad (8.17)$$

где σ^2 — дисперсия вещественной и мнимой составляющих комплексного гауссова шума.

Подставив (8.17) в (8.16), получаем

$$\text{LLR}(\dot{r}_k) = \ln \frac{\frac{1}{2\pi\sigma^2} \exp\left(-|\dot{r}_k - 1|^2 / (2\sigma^2)\right)}{\frac{1}{2\pi\sigma^2} \exp\left(-|\dot{r}_k + 1|^2 / (2\sigma^2)\right)} =$$

$$= \ln \exp\left(-\frac{|\dot{r}_k - 1|^2}{2\sigma^2} + \frac{|\dot{r}_k + 1|^2}{2\sigma^2}\right) = \frac{|\dot{r}_k + 1|^2 - |\dot{r}_k - 1|^2}{2\sigma^2} = \frac{2 \text{Re}(\dot{r}_k)}{\sigma^2}.$$

Таким образом, для случая ФМ-2 мягкие решения пропорциональны вещественной части отсчетов сигнала на выходе СФ приемника.

Постоянный множитель $2/\sigma^2$ не влияет на результат поиска оптимальной кодированной последовательности, поэтому его можно отбросить, что дает правило декодирования мягких решений для случая модуляции ФМ-2:

$$\sum_k a_k \operatorname{Re}(\dot{r}_k) \rightarrow \max_{\{a_k\}}. \quad (8.18)$$

Рассмотрим, как оценивается вероятность ошибки для сверточного кода при использовании ФМ-2 с мягкими решениями. Для этого необходимо воспользоваться формулой (8.12) с соответствующим образом рассчитанной вероятностью перепутывания путей $P(d)$. Определим эту вероятность.

Согласно (8.18), при декодировании мягких решений будут суммироваться вещественные части принимаемого сигнала с коэффициентами $a_k = \pm 1$. Для путей, расположенных на хэмминговом расстоянии d друг от друга, разность математических ожиданий для сумм, накопленных при расчете статистики (8.18) для правильной и неправильной последовательностей, будет равна $2d$, что в d раз больше, чем при рассмотрении одного символа ФМ-2 (расстояние между точками его созвездия равно 2). Так как шум белый, шумовые отсчеты для разных символов независимы, поэтому при суммировании складываются дисперсии шума. Итак, для путей, находящихся на расстоянии d друг от друга, математическое ожидание разности статистик (8.18) увеличивается в d раз по сравнению со случаем некодированной ФМ-2, а среднеквадратическое отклонение шума возрастает в \sqrt{d} раз. Таким образом, отношение сигнал/шум по уровню возрастает в \sqrt{d} раз; кроме того, нужно учесть уменьшение энергии бит, передаваемых в канале связи, пропорционально скорости кодирования r . В результате из (2.11) получаем, что для мягких решений

$$P(d) = Q\left(\sqrt{\frac{2E_b}{N_0}} rd\right).$$

Многочисленные исследования показали, что энергетический выигрыш при переходе от жестких решений к мягким составляет около 2 дБ.

В случае многопозиционной линейной модуляции логарифмические отношения правдоподобия вычисляются для каждого бита символа отдельно, при этом условные ПВ усредняются по всем возможным значениям всех остальных бит:

$$\begin{aligned}
\text{LLR}(\dot{r}_k, m_b) &= \ln \frac{p(\dot{r}_k | b_{m_b} = 1)}{p(\dot{r}_k | b_{m_b} = 0)} = \ln \frac{\frac{1}{M/2} \sum_{n=1}^{M/2} \frac{1}{2\pi\sigma^2} \exp\left(-|\dot{r}_k - \dot{C}_{1n}|^2 / (2\sigma^2)\right)}{\frac{1}{M/2} \sum_{n=1}^{M/2} \frac{1}{2\pi\sigma^2} \exp\left(-|\dot{r}_k - \dot{C}_{0n}|^2 / (2\sigma^2)\right)} = \\
&= \ln \frac{\sum_{n=1}^{M/2} \exp\left(-|\dot{r}_k - \dot{C}_{1n}|^2 / (2\sigma^2)\right)}{\sum_{n=1}^{M/2} \exp\left(-|\dot{r}_k - \dot{C}_{0n}|^2 / (2\sigma^2)\right)}. \tag{8.19}
\end{aligned}$$

Здесь $m_b = 1, \dots, m$ — номер бита, для которого вычисляется LLR; $b_{m_b} = 0$ или 1 — значения этого бита; $\{\dot{C}_{0n}\}$ и $\{\dot{C}_{1n}\}$, $n = 1, \dots, M/2$ — подмножества точек сигнального созвездия, для которых значение бита $b_{m_b} = 0$ и 1 соответственно.

При получении формулы (8.19) использовано предположение, что отдельные биты символа являются статистически независимыми. Для кодированных сигналов это предположение в общем случае несправедливо, так как биты подчиняются использованному алгоритму кодирования. Поэтому применение рассчитанных указанным образом LLR даст при декодировании результаты, несколько худшие, чем истинное правило МП. Для устранения взаимных зависимостей между отдельными битами одного символа в кодированном сигнале используется, в частности, перемежение кодированного битового потока, вводимое между кодером и модулятором. Такая технология получила название *кодированной модуляции с битовым перемежением* (Bit-Interleaved Coded Modulation, BICM), она особенно эффективна для каналов связи с замираниями [10, 8.8.3].

8.4. Сигнально-кодовые конструкции

Как было отмечено в 8.1.3, при использовании помехоустойчивого кодирования появляются потери, связанные с уменьшением энергии бит, передаваемых по каналу связи, из-за внесения избыточности в сообщение. Поэтому требуется использование мощных кодов (большие кодовые ограничения сверточных кодов или большая длина блочного кода) для возмещения этих потерь и обеспечения достаточного выигрыша от кодирования. Однако если рассматривать модуляцию как часть единого процесса кодирования и выбирать ее совместно с кодом таким образом, чтобы увеличить *евклидово* расстояние между сигналами, которые вероятнее всего могут быть перепутаны, то потери можно легко преодолеть и получить заметный выигрыш от кодирования при относительно простых кодах.

Результат совместной оптимизации используемых кода и модуляции называется *сигнально-кодовой конструкцией* (СКК).

При использовании бинарной модуляции евклидово расстояние между кодированными сигналами взаимно-однозначно связано с хэмминговым расстоянием между соответствующими кодовыми словами или кодированными последовательностями, поэтому понятие СКК имеет смысл только для *многопозиционной* (не бинарной) модуляции. Кроме того, в приемнике могут использоваться *только мягкие решения*, так как жесткие решения предполагают, что демодулятор и декодер работают изолированно друг от друга и, следовательно, о каком-либо выигрыше благодаря совместной оптимизации кода и модуляции не может быть и речи.

Далее будут кратко рассмотрены идеи, лежащие в основе двух вариантов СКК:

- решетчатой кодированной модуляции (TCM);
- многоуровневой кодированной модуляции (MLCM).

8.4.1. Решетчатая кодированная модуляция

Принцип решетчатого кодирования (Trellis-Coded Modulation, TCM) был предложен Унгербоеком в 1982 г. Он заключается в том, что для внесения избыточности используется не повышение символьной скорости (что влечет за собой увеличение ширины спектра сигнала), а увеличение числа точек в сигнальном созвездии (M). При этом код и нумерация точек созвездия совместно оптимизируются так, чтобы максимизировать минимальное евклидово расстояние между получающимися кодированными сигналами.

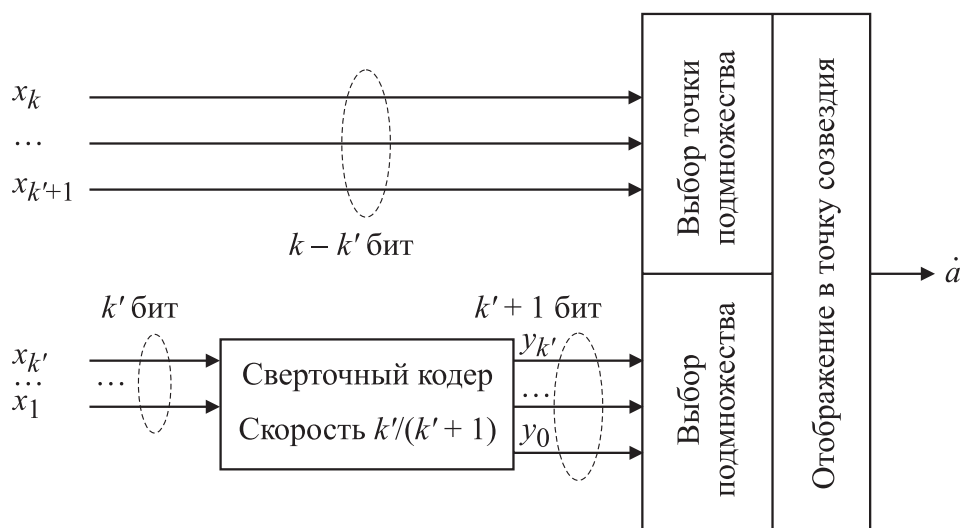


Рис. 8.7. Общая схема решетчатого кодирования

Общая схема решетчатого кодирования представлена на рис. 8.7. Входной битовый поток делится на группы по k бит. Из них $k' < k$ бит ($x_1, \dots, x_{k'}$) подаются на сверточный кодер со скоростью кодирования $k'/(k' + 1)$ (в ре-

зультате кодирования число бит возрастает на единицу), для остальных бит кодирование не используется. Получившаяся группа из $k + 1$ бит используется для выбора точки \dot{a} сигнального созвездия соответствующего размера ($m = k + 1, M = 2^{k+1}$).

Другими словами, созвездие делится на $2^{k'+1}$ подмножеств по $2^{k-k'}$ точек каждое. Кодированные биты $y_0, \dots, y_{k'}$ определяют, какое подмножество используется, а некодированные ($x_{k'+1}, \dots, x_k$) — какая точка из этого подмножества будет выбрана. Отсюда следует, что разделять созвездие на подмножества нужно таким образом, чтобы максимизировать расстояние между ближайшими точками этих подмножеств.

Эффективный метод отображения кодовых бит (символов) в точки сигнального созвездия был предложен Унгербоеком. Метод основан на последовательном разбиении созвездия на две равные части с максимально возможным расстоянием между ближайшими точками. Эта идея продемонстрирована на рис. 8.8 для созвездия ФМ-8.

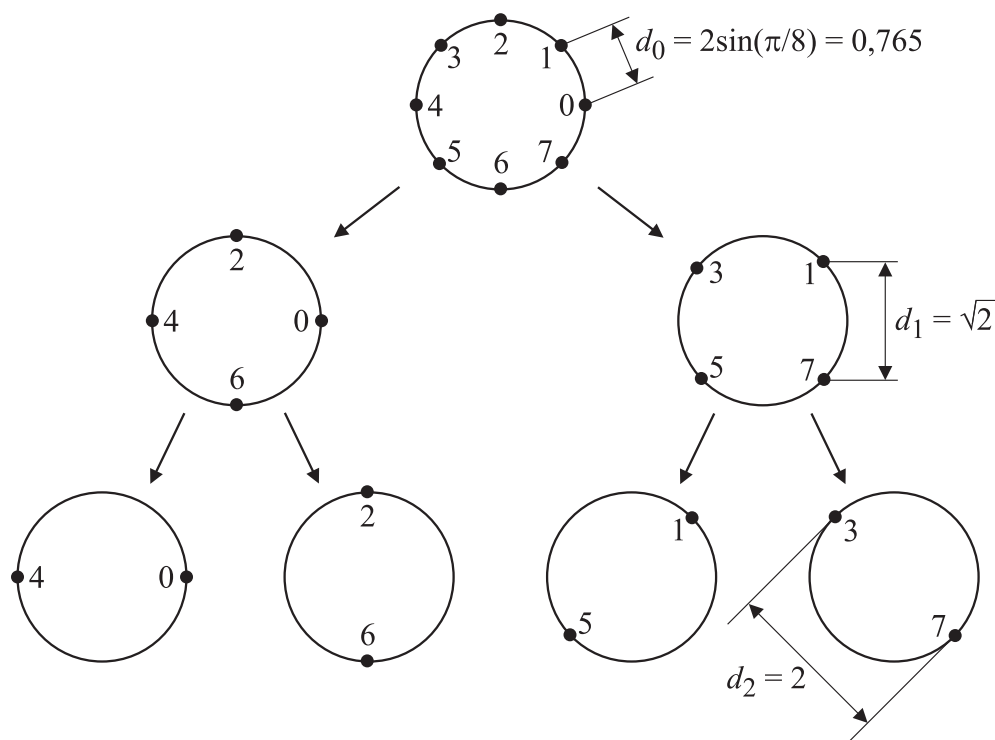


Рис. 8.8. Разбиение Унгербоека для созвездия ФМ-8

Видно, что расстояние между ближайшими точками исходного созвездия ФМ-8 $d_0 = 2 \sin(\pi/8) \approx 0,765$. Разбиение этого созвездия на две половины дает 2 созвездия ФМ-4 с минимальным расстоянием $d_1 = \sqrt{2}$. Наконец, разбиение каждого из этих созвездий на 2 созвездия ФМ-2 дает минимальное расстояние $d_2 = 2$.

Унгербоеком был также разработан эвристический набор правил для назначения символам соответствующих ветвей переходов решетки. Пример такой решетки для четырех внутренних состояний кодера и модуляции ФМ-8 приведен на рис. 8.9.

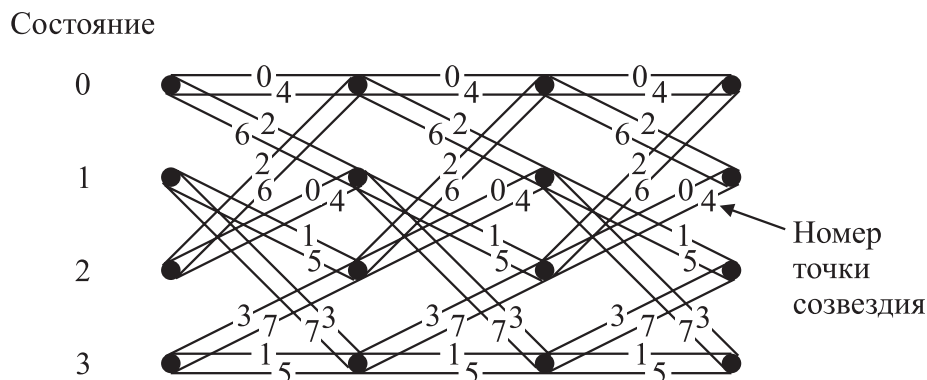


Рис. 8.9. Решетчатая диаграмма ТСМ с четырьмя состояниями для модуляции ФМ-8

«Параллельные» переходы на рис. 8.9 соответствуют разным значениям некодированных бит. Сопоставление рис. 8.9 и 8.8 показывает, что этим парам переходов соответствуют противоположные (наиболее удаленные друг от друга) точки созвездия ФМ-8.

В случае ТСМ (см. рис. 8.7) добавляется один избыточный бит, поэтому сравнивать схему кодирования из рассматриваемого примера нужно с некодированной модуляцией ФМ-4. Определим минимальные евклидовы расстояния между возможными сигналами для этих двух случаев.

При некодированной ФМ-4 это значение равно $\sqrt{2E_g}$ (оно определяется минимальным расстоянием между точками созвездия ФМ-4; E_g — как и раньше, энергия сигнальной посылки). Для рассматриваемого примера ТСМ анализ решетчатой диаграммы (рис. 8.9) показывает, что минимальное расстояние соответствует «параллельным» переходам (т. е. некодированным битам) и равно $2\sqrt{E_g}$. Для пар путей, не содержащих «параллельных» переходов, минимальное евклидово расстояние немного больше (примерно $2,14\sqrt{E_g}$).

Таким образом, данная решетчатая СКК дает асимптотический энергетический выигрыш, равный $20\lg(2/\sqrt{2}) \approx 3$ дБ.

В приемнике зашумленный сигнал декодируется по правилу максимального правдоподобия декодером Витерби с мягким входом (т. е. в качестве минимизируемой метрики используется евклидово расстояние между сигналами).

Аналогичным образом строятся решетчатые СКК и для КАМ-созвездий.

В практических системах, чтобы упростить процедуру обработки символов на приемной стороне, обычно применяется дифференциальное кодирование. Это позволяет получить СКК, инвариантные к некоторым сдвигам фазы. Необходимые для этого коды обычно оказываются нелинейными.

Схемы решетчатой кодированной модуляции были также разработаны для *многомерных символов*. Многомерные символы представляют собой последовательность одномерных (АМ) или двумерных (КАМ) символов. С помощью многомерных ТСМ можно получить такой же выигрыш в свободном расстоянии, как и в случае двумерных символов, при использовании двумерных сигнальных созвездий с меньшим числом точек.

8.4.2. Многоуровневая кодированная модуляция

Многоуровневая кодированная модуляция (MultiLevel Coded Modulation, MLCM) — это еще один способ объединения процессов модуляции и кодирования. Данный метод был предложен в 1977 г. Имаи и Хиракавой, краткая информация о нем имеется в [9].

Подобно ТСМ, MLCM также использует иерархическое разбиение сигнального созвездия. Однако, если при использовании ТСМ часть бит многобитового символа кодируется помехоустойчивым кодом, а часть (возможно) остается некодированной, то сущность MLCM состоит в том, что каждый отдельный бит символа кодируется *своим собственным* кодом.

Таким образом, сигнальное созвездие, содержащее $M = 2^m$ точек, иерархически (m раз) разбивается на 2 подмножества. Выбор одного из двух подмножеств на каждом уровне осуществляется в соответствии со значениями бит кодовых слов (или, в случае сверточного кодирования, кодовых последовательностей) *компонентных кодов*. Всего необходимо m компонентных кодов, каждый из которых соответствует своему уровню иерархии.

ТСМ может трактоваться как частный случай MLCM, для этого кодированные биты символа рассматриваются как единый символ недвоичного компонентного кода, а некодированные биты трактуются как символы других компонентных кодов, имеющих нулевую избыточность, т. е. скорость кодирования, равную единице.

Одним из преимуществ MLCM является гибкость возможностей согласования минимальных евклидовых расстояний подмножеств сигнального созвездия на каждом уровне разбиения с расстояниями Хэмминга компонентных кодов. За счет этого и удастся увеличить минимальное евклидово расстояние между кодированными сигналами в СКК данного типа.

Структурная схема кодера MLCM показана на рис. 8.10.

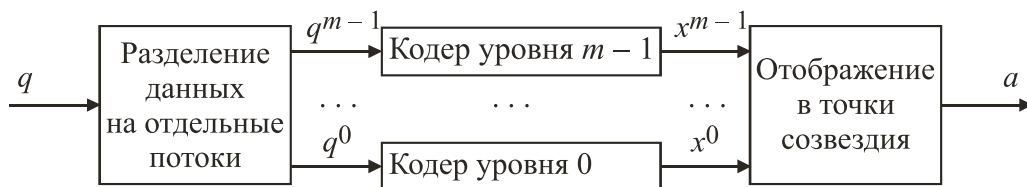


Рис. 8.10. Кодер MLCM

Входной битовый поток q разделяется на m подпотоков, каждый из которых кодируется своим помехоустойчивым кодом. Биты (m штук), получаемые в каждый момент на выходе всех кодеров, образуют m -битовый символ, который отображается в точку a используемого сигнального созвездия.

Возможным способом декодирования MLCM является *каскадный декодер*, структурная схема которого показана на рис. 8.11.

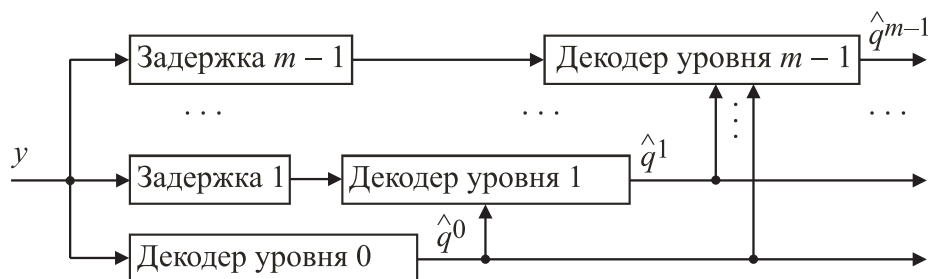


Рис. 8.11. Каскадный декодер MLCM

Отдельные битовые подпотоки декодируются своими *отдельными* декодерами, при этом каждый декодер использует *оценки* бит, выданные декодерами более низких уровней (для этого необходимо использование соответствующих задержек), и считает, что на более высоких уровнях кодирование *не применяется*.

Следует отметить, что каскадный декодер не является МП-декодером для данной СКК в целом. Однако показано, что при его использовании в сочетании с должным выбором параметров компонентных кодов в пределе обеспечивается приближение к пропускной способности канала.

СПИСОК ЛИТЕРАТУРЫ

1. Скляр Б. Цифровая связь. Теоретические основы и практическое применение / Пер. с англ. — М.: Издательский дом «Вильямс», 2003.
2. Прокис Дж. Цифровая связь / Пер. с англ.; Под ред. Д. Д. Кловского. — М.: Радио и связь, 2000.
3. Ипатов В. П. Широкополосные системы и кодовое разделение сигналов. Принципы и приложения / Пер. с англ. — М.: Техносфера, 2007.
4. Иванов М. Т., Сергиенко А. Б., Ушаков В. Н. Теоретические основы радиотехники: Учеб. пособие / Под ред. В. Н. Ушакова. — 2-е изд. — М.: Высш. шк., 2008.
5. Сергиенко А. Б. Цифровая обработка сигналов: Учеб. пособие. — 3-е изд. — СПб.: БХВ-Петербург, 2011.
6. Радиотехнические системы: Учеб. для студ. высш. учеб. заведений / Под ред. Ю. М. Казаринова. — М.: Изд. центр «Академия», 2008.
7. Соколов А. И., Юрченко Ю. С. Радиоавтоматика: Учеб. пособие для вузов. — М.: Изд. центр «Академия», 2011.
8. Вернер М. Основы кодирования. — М.: Техносфера, 2006.
9. Морелос-Сарагоса Р. Искусство помехоустойчивого кодирования. Методы, алгоритмы, применение / Пер. с англ. — М.: Техносфера, 2005.
10. Голдсмит А. Беспроводные коммуникации / Пер. с англ. — М.: Техносфера, 2011.

Оглавление

1. ОБЩАЯ СТРУКТУРА И ОСНОВНЫЕ ХАРАКТЕРИСТИКИ СИСТЕМ ЦИФРОВОЙ СВЯЗИ.....	3
1.1. Понятие цифровой связи и ее краткая история.....	3
1.2. Общая структура системы цифровой связи	4
1.3. Специализированные и универсальные системы связи	5
1.4. Важнейшие параметры систем цифровой связи	6
1.5. Скремблирование	9
1.6. Алгоритм Витерби	11
1.7. АБГШ-канал и его пропускная способность.....	14
2. МОДУЛЯЦИЯ	16
2.1. Линейная цифровая модуляция	16
2.2. Нелинейная цифровая модуляция	33
2.3. Эффективность некодированной модуляции	46
3. МЕЖСИМВОЛЬНАЯ ИНТЕРФЕРЕНЦИЯ	48
3.1. Математическая модель канала с МСИ	48
3.2. Оптимальный алгоритм приема сигнала (алгоритм Витерби)	49
3.3. Оптимальный фильтр Винера	55
3.4. Линейный эквалайзер	58
3.5. Эквалайзер с обратной связью по решению.....	62
3.6. Многочастотная модуляция (технологии OFDM, DMT)	64
4. ЗАМИРАНИЯ И РАЗНЕСЕНИЕ	74
4.1. Энергетические соотношения в радиолинии	74
4.2. Причины возникновения и общие характеристики замираний.....	76
4.3. Математическая модель канала с замираниями.....	77
4.4. Вероятностные свойства коэффициента передачи канала связи	78
4.5. Корреляционные и спектральные характеристики коэффициента передачи канала связи	80
4.6. Оценка помехоустойчивости различных видов модуляции в рэлеевском канале связи	82
4.7. Понятие вероятности отказа связи	84
4.8. Идея разнесения. Способы организации ветвей разнесенного приема....	85
4.9. Способы обработки принятых сигналов при разнесенном приеме	85
4.10. Перемежение как способ борьбы с группированием ошибок	90
5. РАСШИРЕНИЕ СПЕКТРА	92
5.1. Общие принципы расширения спектра	92
5.2. Прямое расширение спектра	97
5.3. Псевдослучайная перестройка рабочей частоты	106

6. МНОЖЕСТВЕННЫЙ ДОСТУП.....	112
6.1. Структуры многопользовательских систем	112
6.2. Разделение каналов (множественный доступ)	113
6.3. Случайный доступ.....	118
7. СИНХРОНИЗАЦИЯ	120
7.1. Задачи синхронизации	120
7.2. Ухудшение помехоустойчивости из-за ошибок синхронизации	121
7.3. Классификация алгоритмов синхронизации	125
7.4. Фазовая синхронизация	127
7.5. Временная синхронизация	133
8. ПОМЕХОУСТОЙЧИВОЕ КОДИРОВАНИЕ.....	138
8.1. Основные принципы и термины.....	138
8.2. Блочные коды.....	141
8.3. Сверточные коды.....	149
8.4. Сигнально-кодовые конструкции.....	156
СПИСОК ЛИТЕРАТУРЫ.....	162

Сергиенко Александр Борисович

Цифровая связь

Учебное пособие

Редактор Э. К. Долгатов

Подписано в печать 26.06.12. Формат 60×84 1/16. Бумага офсетная.
Печать офсетная. Гарнитура «Times New Roman». Печ. л. 10,25.
Тираж 50 экз. Заказ 76.

Издательство СПбГЭТУ «ЛЭТИ»
197376, С.-Петербург, ул. Проф. Попова, 5